9-1-2015 12:00 AM

Inferring plastid metabolic pathways within the nonphotosynthetic free-living green algal genus Polytomella

Sara Asmail
The University of Western Ontario

Supervisor
Dr. David Smith
The University of Western Ontario

Graduate Program in Biology
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

INFERRING PLASTID METABOLIC PATHWAYS WITHIN THE NONPHOTOSYNTHETIC FREE-LIVING GREEN ALGAL GENUS *POLYTOMELLA*

(Thesis format: Integrated Article)

by

Sara R. Asmail

Graduate Program in Biology

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

The advent of photosynthesis facilitated the evolution of aerobic life on Earth. However, species such as *Prototheca wickerhamii* and *Plasmodium falciparum*, among many others, have lost photosynthesis and opted for a free-living/parasitic lifestyle. Despite this loss, these species have retained the plastid for its metabolic pathways, without which they would die. *Polytomella* is a nonphotosynthetic free-living alga, closely related to the photosynthetic model organism *Chlamydomonas reinhardtii*, and has been shown to lack a plastid genome. I set out to determine *Polytomella* plastid metabolic pathways using bioinformatics to look for mRNA and DNA homologous sequences matching pathway enzymes in model organisms. Interestingly, *Polytomella* possesses all enzymes required for amino acid, tetrapyrrole, starch, and carotenoid biosynthesis. However, it lacks enzymes that are plastid-encoded and/or are involved in chlorophyll biosynthesis. The progressive loss of carotenoid biosynthesis provides insight into photosynthetic loss, and *Polytomella* shows higher plastid functional complexity in comparison to parasitic species.

Keywords

Algae; *Chlamydomonas*; chloroplast; organelle; plastid; photosynthesis; metabolic pathways; starch; carotenoids; amino acids; biosynthesis; parasites

Co-Authorship Statement

Appendix A—Next generation sequencing data suggest that certain nonphotosynthetic green plants have lost their plastid genomes.

This is a letter to New Phytologist that was coauthored by Dr. David Smith, my supervisor, and myself. I contributed intellectually to the writing of this letter while my supervisor wrote and polished the final manuscript, and submitted it to the journal.

Appendix B—Retention, erosion, and loss of the carotenoid biosynthetic pathway in the nonphotosynthetic green algal genus *Polytomella*.

This was again a letter to the journal New Phytologist that was coauthored by myself and Dr. David Smith. I contributed intellectually, as well as conducted the data analyses and wrote the draft manuscript. Dr. Smith verified my results and wrote, polished, and submitted the final manuscript to the journal.

Both articles are free-access papers, and further publication permitted.

Acknowledgments

I would like to thank my supervisor, Dr. David Smith, for all his invaluable help and patient guidance throughout my graduate years. It was truly a blessing to have a supervisor like him. He was always very comfortable to talk to and was always eager to provide wonderful advice and suggestions, whenever and wherever I had trouble with the project. It is a great feeling to know that you could always rely on someone to debate your points of view as well as to offer help. In my graduate experience with David, I've learnt that the key to succeeding in science requires patience, persistence, and hardwork, and that no organism is too small or too insignificant to inquire about and investigate. Thank you, David, for always being there and for helping me to grow up as a scientist and succeed with this milestone.

I would also like to thank my advisory committee, Dr. Norm Hüner and Dr. Kathleen Hill, for all their great assistance and support. Their questions and suggestions made for a valuable project, and helped me think of new directions. They have also facilitated my realization that scientific thinking requires investigating the whole organism, and all of its genetic and environmental aspects, and knowing that one could spend a lifetime trying to know one organism.

I would like to thank my lab mates and office colleagues, in specific, Michael Del Vasto and Jennifer McDonald, for all their stress relieving techniques and for always being there to help me emotionally and academically.

Finally, I would like to thank my family and especially my mother for always being a great role model and showing me that courage and kindness do not require strength, power, or money; that you only need to do the right thing. You've taught me that even ordinary people can be heroes.

Table of Contents

Abstract ii

Co-Authorship Statement iii

Acknowledgments iv

Table of Contents v

List of Tables viii

List of Figures ix

List of Appendices x

List of Abbreviations xi

Chapter 1 1

1 The evolutionary loss of photosynthesis: from parasites to *Polytomella* 1

1.1 Photosynthesis 1

1.2 The origins and spread of photosynthesis 2

1.2.1 Cyanobacteria to Eukarya 2

1.2.2 Integration and the rise of Archaeplastida 3

1.2.3 Plastid Evolutionary Origins 5

1.3 The Loss of Photosynthesis 5

1.3.1 The general loss of photosynthesis 5

1.3.2 Loss of photosynthesis yet plastid retention hypotheses 9

1.4 Metabolic Pathways in Nonphotosynthetic Species 10

1.5 *Polytomella* 12

1.6 Thesis Rationale and Hypothesis 14

1.7 References 16

Chapter 2 22

2 The Starch Metabolic Pathway 22

2.1 Introduction: starch metabolism at a glance. 22
2.2 Starch biosynthesis pathway in detail. 22
2.3 Methods: the transcriptomic and genomic data analyses of *Polytomella* spp. 24
2.3.1 Using *C. reinhardtii* as a reference organism 24
2.3.2 Transcriptomic analyses 25
2.3.3 Genomic Analyses 28
2.4 Results 29
2.5 Discussion 30
2.6 Abbreviations 32
2.7 References 33
Chapter 3 37
3 Tetrapyrrole Biosynthetic Pathway 37
3.1 Introduction: tetrapyrrole biosynthesis at a glance. 37
3.2 Tetrapyrrole biosynthetic pathway in detail. 38
3.3 Methods 40
3.4 Results 40
3.5 Discussion 43
3.6 Abbreviations 44
3.7 References 46
Chapter 4 49
4 Amino Acids Biosynthetic Pathway 49
4.1 Introduction: Amino acid pathways at a glance. 49
4.2 Amino acid pathways in detail. 53
4.3 Methods 62
4.4 Results 62
4.5 Discussion 67

4.6 Abbreviations 70
4.7 References 74
Chapter 5 78
5 The Carotenoid Biosynthetic Pathway 78
5.1 Introduction: carotenoids at a glance. 78
5.2 Carotenoids in detail. 80
5.3 Methods 83
5.4 Results 83
5.5 Discussion 84
5.6 Abbreviations 88
5.7 References 90
Chapter 6 94
6 Plastid-localized metabolic pathways and plastid complexity in *Polytomella* spp. 94
6.1 A summary of *Polytomella* plastid metabolic pathways 94
6.2 *Polytomella* spp. plastid complexity 95
6.3 Concluding remarks 96
6.4 References 99
Appendices 100

List of Tables

Table 1: Examples of nonphotosynthetic green algae and land plants including a description of their plastid genomes.....7

Table 2: Plastid metabolic pathways within nonphotosynthetic algae of primary or secondary endosymbiotic plastid origins.....12

Table 3: *Polytomella* RNA-Seq data summary, also shown in Figures 4 and 5. 26

Table 4: Starch metabolic pathway enzymes, shown in order of appearance within the pathway.....29

Table 5: Tetrapyrrole biosynthetic pathway enzymes.....42

Table 6: Ammonia assimilation enzymes.....63

Table 7: Aromatic amino acid enzymes. 64

Table 8: Aspartate-derived amino acid enzymes.....65

Table 9: Branched-chain amino acid biosynthetic enzymes. 66

Table 10: Arginine and proline biosynthetic pathway. 66

Table 11: Carotenoid biosynthesis enzymes. 86

Table 12: Plastid metabolic pathways within nonphotosynthetic algae and *Polytomella* spp.95

Table 13: Metabolic pathways in *C. reinhardtii*, *Helicosporidium* sp., *Polytomella* spp. and *P. wickerhamii*, respectively.....97

List of Figures

Figure 1: Outline of different plastid origins in primary and secondary endosymbioses lineages.. 4

Figure 2: Tree of photosynthetic and nonphotosynthetic green algae, including *Polytomella* species....13

Figure 3: Consensus starch metabolic pathway in green algae. 23

Figure 4: Outline of the DNA and RNA sequencing project for *Polytomella* spp. 27

Figure 5: Data analyses summary for the four *Polytomella* spp. 28

Figure 6: Tetrapyrrole biosynthetic pathway. 39

Figure 7: Summary of amino acid biosynthetic pathways predicted to occur within the *Polytomella* plastid. 52

Figure 8: Aromatic amino acid biosynthetic pathway. 55

Figure 9: Aspartate-derived amino acid biosynthetic pathway. 57

Figure 10: Branched chain amino acid biosynthetic pathway. 58

Figure 11: Arginine and proline biosynthetic pathway. 61

Figure 12: The carotenoid biosynthetic pathway. 82

List of Appendices

Appendix A: Next-generation sequencing data suggest that certain nonphotosynthetic green plants have lost their plastid genomes. 100

Appendix B: Retention, erosion, and loss of the carotenoid biosynthetic pathway in the nonphotosynthetic green algal genus *Polytomella*. Paper submission to *New Phytologist*, along with all of its supplementary data, tables, and figures. 105

List of Abbreviations

This is a comprehensive list of all the enzymes for each specific pathway investigated in this thesis, along with their abbreviations and GenBank accession or E.C. number.

Enzyme	Abbreviation	E.C. or Acc. no.
Starch Biosynthesis		
Plastidial phosphoglucomutase	GPM or PGM	5.4.2.2
ADP-glucose pyrophosphorylase	AGPase	2.7.7.27
Starch synthase	SS	2.4.1.21
Branching enzyme	BE	2.4.1.18
Debranching enzyme	DBE	3.2.1.68
α-1,4-glucanotransferases	GT	2.4.1.25
Starch phosphorylases	SP	2.4.1.1
Maltose exporter like protein	MEX1	EDP05341
Glucan-water dikinases	GWD	2.7.9.5
α-amylases	AMS	3.2.1.1
β-amylases	BAMS	3.2.1.2
Tetrapyrrole Biosynthesis		
Glutamyl-tRNA synthetase	GTS	6.1.1.17
Glutamyl-tRNA reductase	GTR	1.2.1.70
Glutamate-1-semialdehyde aminotransferase	GSAT	5.4.3.8
Porphobilinogen synthase, ALA dehydratase	PBGS/ALAD	4.2.1.24
Hydroxymethylbilane synthase	PBGD	4.3.1.8
Uroporphyrinogen III synthase	UROS	4.2.1.75
Uroporphyrinogen III decarboxylase	UROD	4.1.1.37
Coproporphyrinogen III oxidative decarboxylase	CPX	1.3.3.3

Protoporphyrinogen IX oxidase	PPX	1.3.3.4
Mg chelatase	CHLD	EDP07156
	CHLH1	EDP07149
	CHLH2	EDP00299
	CHLI1	EDP04965
	CHLI2	EDP05319
Mg-chelatase activating protein	GUN4	EDP07013
Mg-protoporphyrin IX methyltransferase	PPMT	2.1.1.11
Mg-protoporphyrin IX monomethylester cyclase	CRD1	EDP04035
	CTH1	EDP05493
Light-dependent protochlorophyllide oxidoreductase	POR	1.3.1.33
Light-independent protochlorophyllide reductase	CHLB	AAB06265
	CHLL	CAA43020
	CHLN	AAN41268
Divinyl-protochlorophyllide vinyl reductase	DVR	1.3.1.75
Chlorophyll synthase	CHS/CHLG	2.5.1.62
Chlorophyllide a oxidase	CAO	EDP09913
Geranylgeranyl reductase	GGR	EDP09986
Chlorophyllase	CLH	3.1.1.14
Precorrin-2-synthase	UPM	2.1.1.107
Precorrin-2-dehydrogenase		1.3.1.76
Sirohydrochlorin ferrochelatase	SIRB	4.99.1.4
Ferrochelatase	FeC	4.99.1.1
Heme-c-ligase		4.4.1.17
Heme oxygenase	HMOX1	EDP06362
	HMOX2	EDO98578
Aromatic Amino Acids		
Aromatic amino acid hydroxylase	AAH1	EDP09772
Arogenate dehydratase	ADT1	EDP08558

Arogenate/prephenate dehydrogenase	AGD1	EDP01548
Anthranilate synthase, α-subunit	ASα/ ANS1	EDP05111
Anthranilate synthase, β-subunit	ASβ/ ASB1	EDP09587
Phosphoribosylanthranilate isomerase	PAI/ ASB2	EDP00638
Aspartate aminotransferase	AST1/ AspAT	EDP02192
	AST2	EDP09991
	AST3	EDP07735
	AST4	EDP08586
	AST5	EDO99564
Chorismate mutase	CHM1	EDP00124
3-Dehydroquinate synthase	DHQS	EDP01025
Indole-3-glycerol-phosphate synthase	IGS/ IGPS/ MAA4	EDP09023
Tryptophan synthase, β-subunit	TSB/ MAA7	EDP06079
Anthranilate phosphoribosyltransferase	APRT/ PRT1	EDP06478
3-deoxy-D-arabino-heptylosonate 7-phosphate synthase	DAHPS/ SHKA1	EDP00515
Bifunctional dehydroquinate dehydratase-shikimate: NADP oxidoreductase	SHKD1/ DHQase-SORase	EDP02779
Shikimate kinase	SHKF1	EDP05661
5-enolpyruvylshikimate-3-phosphate synthase	EPSPS/SHKG1	EDO96795
Chorismate synthase	CS/ SHKH1	EDP01398
Tryptophan synthetase, α-subunit	TSA/ MAA1	EDP00665
Arginine and Proline Biosynthesis		
N-acetyl-L-glutamate kinase	NAGK/ AGK1	EDP09199
Argininosuccinate synthase	AS/ AGS1	EDP00857
N-acetylornithine deacetylase	NAOD/ AOD1	EDP07452
N-acetyl-γ-glutamyl-phosphate reductase	ARG1	EDP01479
Argininosuccinate lyase	ASL/ ARG7	EDP09253
N-acetylornithine aminotransferase	NAOAT/ ARG9	EDO99676
Carbamoyl phosphate synthase, large subunit	CMPL1/ CPS	EDO96236

Carbamoyl phosphate synthase, small subunit	CMPS1	EDP05092
Δ^1-pyrroline-5-carboxylate synthetase	GSD1	EDP01490
N-acetylglutamate synthase, monofunctional	LCI8	EDP02142
N-acetylglutamate synthase, bifunctional	NAGS	EDO98751
Ornithine-δ-aminotransferase	OAT/ OTA	EDO097297
Ornithine carbamoyltransferase	OTC	EDP05375
Δ^1-pyrroline-5-carboxylate reductase	P5CR/ PCR1	EDP00128/ EDP00129
γ-glutamyl kinase	PROB1	EDP04671
	PROB2	EDP04670
Branched chain amino acid biosynthetic enzymes		
Dihydroxyacid dehydratase	AAD1	EDP03205
Acetohydroxy acid isomeroreductase	AAI1	EDP06428
Acetolactate synthase, large subunit	ALSL1	EDP01876
Acetolactate synthase, small subunit	ALSS1	EDO98300
Branched chain aminotransferase	BCAT1	EDP06865
	BCAT2	EDP07184
	BCAT3	EDP08065
3-isopropylmalate dehydratase, large subunit	LEU1L	EDO97224
Isopropylmalate dehydratase, small subunit	LEU1S	EDP08379
2-isopropylmalate synthase	LEU2	EDP08580
3-isopropylmalate dehydrogenase	LEU3	EDP07327, EDP07328
Threonine deaminase	THD	EDP06791
Aspartate derived amino acid biosynthetic enzymes		
Bifunctional aspartate kinase/homoserine dehydrogenase	AHD1	EDP01964
Monofunctional aspartate kinase	ASK1	EDP08069
Asparagine synthetase	ASNS	EDO97202

Aspartate semialdehyde dehydrogenase	ASSD	EDP02211
Cystathionine γ-synthase	CGS1	EDP01529
Diaminopimelate epimerase	DAE1	EDO99946
L,L-diaminopimelate aminotransferase	DPA1	EDP03630
Diaminopimelate decarboxylase	DPD1	EDP09979
Dihydrodipicolinate synthase	DPS1	EDP07434
Dihydrodipicolinate reductase	DPR1	1.3.1.26
Homoserine dehydrogenase	HSD1	EDP07408
Homoserine kinase	HSK1	EDP06874
Cystathionine β-lyase	METC	EDO98869
Cobalamin-independent methionine synthase	METE/ MS	EDO96787
Cobalamin-dependent methionine synthase	METH1/ MS	EDP08397
Threonine synthase	THS1	EDP08010
NH_3 assimilation and glutamate/glutamine biosynthesis enzymes		
Glutamate dehydrogenase	GDH1	EDP02540
	GDH2	EDO97071
Glutamine synthetase	GLN1/ GS1	EDP07598
	GLN2/ GS2	EDP03611
	GLN3/ GS2	EDP03496
	GLN4/ GS1	EDP03343
Ferrodoxin-dependent glutamate synthase	GSF1/ Fd-GOGAT	EDO96739
NADH-dependent glutamate synthase	GSN1/ NADH-GOGAT	EDP03651
Carotenoid Biosynthetic Pathway		
1-deoxy-D-xylulose-5-phosphate synthase	DXS	CAA07554
1-deoxy-D-xylulose-5-phosphate reductoisomerase	DXR	EDP02894
4-diphosphocytidyl-2-C-methyl-D-erythritol synthase	CMS	EDO99224
4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	CMK	EDP02028
2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase	MCS	EDP05431

4-hydroxy-3-methylbut-2-enyl-1-diphosphate synthase	HDS	AAX54905
Isopentenyl/ dimethylallyl diphosphate synthase	IDS	EDO97597
Isopentenyl/ dimethylallyl diphosphate delta isomerase	IDI	AAC32601
Geranylgeranyl diphosphate synthase	GGPPS	EDO96545
Phytoene synthase	PSY	AAT38474
Phytoene desaturase	PDS	AAT38476
ζ-Carotene desaturase	ZDS	EDP07040
Carotenoid isomerase	CRTISO	EDO99516
Lycopene-β-cyclase	LCYB	AAX54906
Lycopene-ε-cyclase	LCYE	AAT46065
Carotene-β-hydroxylase	CHYB	AAX54907
Carotene-β-hydroxylase	LUT5	EDO98047
Carotene-ε-hydroxylase	LUT1	ABQ59243
Zeaxanthin epoxidase	ZEP	AAO34404
Violaxanthin de-epoxidase	VDE	EDP02194.1
Neoxanthin synthase	NSY	--
Loroxanthin synthase	LSY	--
Carotene-β-ketolase	BKT	AAX54908

Chapter 1

1 The evolutionary loss of photosynthesis: from parasites to *Polytomella*

1.1 Photosynthesis

Photosynthesis is one of the processes that have made Earth habitable. It is more than three billion years old and is a driving force of biological evolution and human existence (Gould *et al.* 2008; Sakamoto *et al.* 2008). In its most basic form, photosynthesis converts light, water, and carbon dioxide (produced via cellular respiration and fossil combustion) into sugars and oxygen. Photosynthesis is crucial for life: it is the only known biochemical means of oxygen production and provides organic carbon molecules to photosynthetic organisms (Bendall *et al.* 2008). Not only is photosynthesis beneficial to photosynthesizing species, it is also pivotal to the survival and diversity of nonphotosynthetic organisms. Photosynthetic organisms undertake important roles as food resources, ecological and ecosystem stabilizers, and air purifiers. Moreover, one could think of photosynthesis as the prototype to current technological advances such as solar panels that promise to decrease human reliance on fossil fuels and lower the extent of climate change (Bullis 2013; Savage 2013).

Although photosynthesis is often associated with land plants, much of the known photosynthetic diversity is microbial (Fuller *et al.* 2006). In fact, marine algae (phytoplankton) are responsible for 45-50% of global photosynthetic oxygen production (Falkowski 1994; Field *et al.* 1998; Roach 2004). Photosynthetic species populate almost every ecosystem on Earth; they are found in some of the most remote and diverse lineages on the tree of life (Archibald 2009; Hoefnagels 2011), and new photosynthetic groups and species are still being uncovered (Kim *et al.* 2011; Janouškovec *et al.* 2010).

But sometimes the lights go out; sometimes species lose their photosynthetic capabilities. Throughout the eukaryotic domain there are numerous examples of algae and land plants that can no longer derive energy from sunlight and have thus reverted to a heterotrophic state (Table 1; Keeling 2010; Wicke *et al.* 2013; Molina *et al.* 2014; Smith and Lee 2014;

Tartar and Boucias 2004). Some of these photosynthetic burnouts are parasites, others are free living, most still retain some of the machinery that was once used to carry out photosynthesis (plastids), and they all have a lot to teach us about the origins and evolution of photosynthesis (Wicke *et al.* 2013; Molina *et al.* 2014; Smith and Lee 2014; Tartar and Boucias 2004; de Koning and Keeling 2006; Borza *et al.* 2005). Recent genome and transcriptome sequencing of free-living, parasitic, and nonphotosynthetic green algae are changing, and sometimes reaffirming, long-held views about plastid evolution and diversity. These studies have provided insights into plastid genome retention and loss (Smith and Lee 2014) and revealed unique plastid metabolic pathways that are unrelated to photosynthesis (Pombert *et al.* 2014). Green algae are quickly becoming the leading systems for comparative genomics and evolutionary studies of photosynthetic and nonphotosynthetic species.

1.2 The origins and spread of photosynthesis

1.2.1 Cyanobacteria to Eukarya

Eukaryotic photosynthesis first emerged ~1.5 billion years ago through the endosymbiosis of a cyanobacterium by a heterotrophic eukaryote (Archibald 2009; Keeling 2010; Hedges *et al.* 2004; Yoon *et al.* 2004). Over time, the cyanobacterial endosymbiont yielded its autonomy and genetic and biochemical control to the eukaryotic host, eventually becoming a *bona fide* photosynthetic organelle, which we call a chloroplast or plastid (Hedges *et al.* 2004; Yoon *et al.* 2004; Archibald 2009; Keeling 2010). This transition from endosymbiont to organelle occurred over millions of years and involved massive and ongoing migrations of genes from the cyanobacterial genome to the host nuclear genome through a process known as endosymbiotic gene transfer (Gould *et al.* 2008). The transfer of genes may have occurred due to the more complex regulatory machinery of nuclear genes in comparison to organelle genomes and the lower rate of mutations in the nucleus (Herrmann 1997; Martin and Herrmann 1998). Consequently, the genomes of present-day plastids are highly reduced relative to those of extant, free-living cyanobacteria, and the majority of plastid proteins are encoded by nuclear DNA (nucDNA) and imported (post-translationally) into the plastid via various import machineries (Keegstra and Cline 1999; Gould *et al.* 2008; Soll and Schleif 2004;

Hormann *et al.* 2007). Only certain essential anabolic and housekeeping genes are still located in the plastid DNA (ptDNA) (de Koning and Keeling 2004). These genes include ones involved in plastid-genome expression such as genes encoding tRNAs, RNA polymerases, ribosomal RNAs, clpP, ycf1, and ycf2 (Barkan 2011; Smith and Lee 2014).

1.2.2 Integration and the rise of Archaeplastida

Eukaryotic plastids can have very different evolutionary histories. Primary plastids descend directly from the endosymbiosis of a cyanobacterium (Archibald 2009; Keeling 2010), and are restricted to the Archaeplastida (Plantae *sensu lato*), which is one of approximately five supergroups that make up the eukaryotic domain (Keeling *et al.* 2005). The Archaeplastida comprises glaucophytes, red algae (rhodophytes), green algae, and land plants (Figure 1; Adl *et al.* 2005). The latter two lineages form the Viridiplantae, and are often referred to as "the Green Lineage." Viridiplantae can be further subdivided into the Streptophyta—which consists of land plants and the green algae most closely related to land plants (charophyte algae)—and the Chlorophyta (what most people refer to as green algae) (Picket-Heaps and Marchant 1972; Bremer 1985; Lemieux *et al.* 2007).

Once established within the Archaeplastida, plastids and photosynthesis spread laterally across the eukaryotic tree via secondary, tertiary, and serial endosymbioses, all of which involve the engulfment of a photosynthetic eukaryote by another eukaryote (Bhattacharya *et al.* 2004; Gould *et al.* 2008). A surprisingly large number of lineages harbor eukaryote-eukaryote-derived plastids. For instance, photosynthetic euglenids, chlorarachniophytes, and some dinoflagellate algae acquired their plastids by engulfing green algae, whereas apicomplexan parasites, stramenopile algae, and haptophytes have plastids from red algae (Figure 1; Gilson *et al.* 2006; Archibald 2007; Douglas *et al.* 1999; Sanchez-Puerta *et al.* 2007; Yoon *et al.* 2002; Funes *et al.* 2002; Waller *et al.* 2003). The types of species with eukaryote-eukaryote-derived plastids can be remarkably diverse. For instance, apicomplexans are associated with serious diseases, such as malaria and toxoplasmosis, dinoflagellate algae are famous for forming massive blooms and "red tides," and euglenids are renowned for their wide-ranging forms of nutrition, from free-living to parasitism to bacteriovery to photoautotrophy (Wilson *et al.* 1996; Snounou *et al.* 1993;

Anderson 1994; Leander *et al.* 2001). Things get even more confusing when considering species whose plastids derive from tertiary and serial endosymbiosis events.

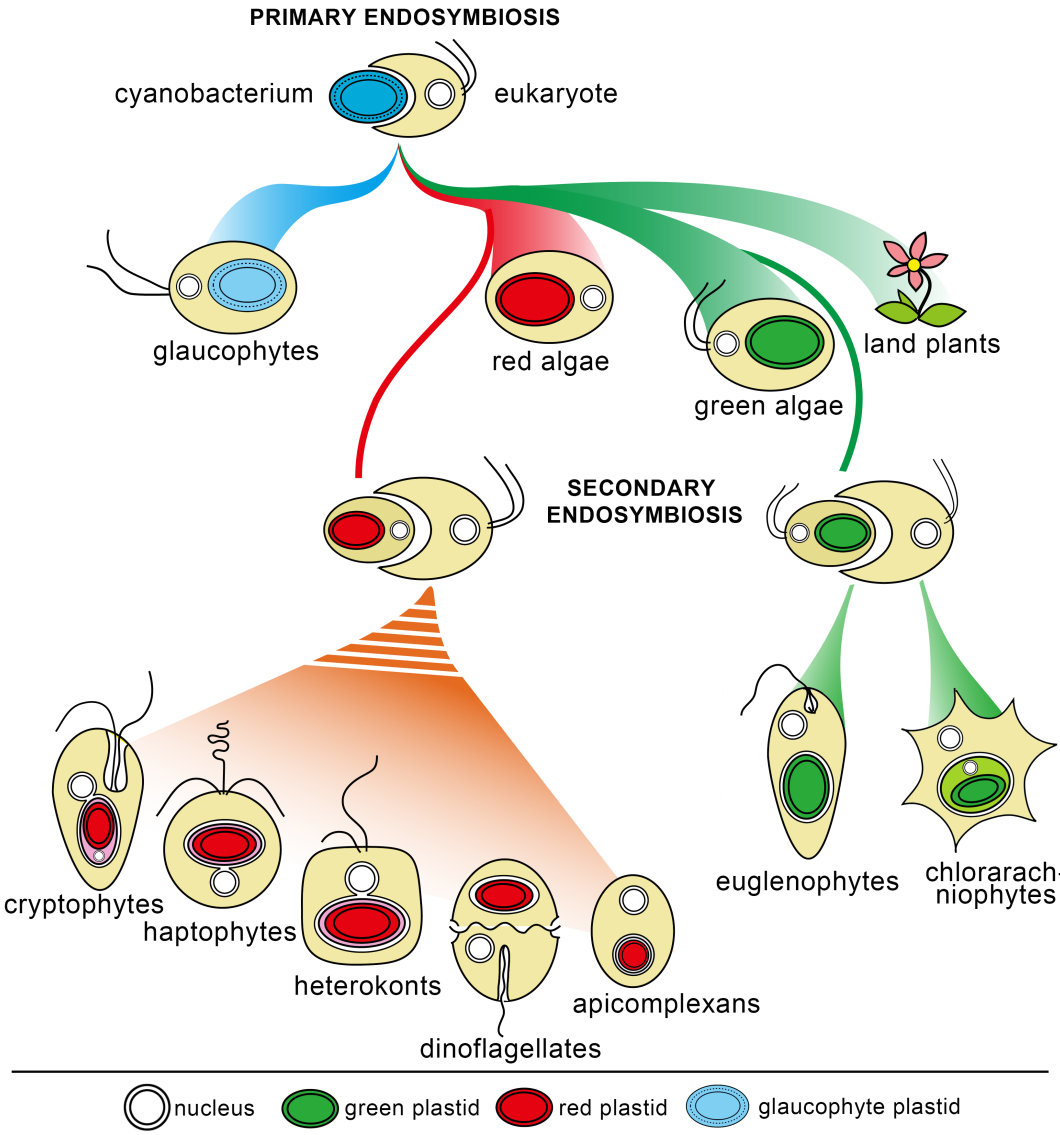


Figure 1: Outline of different plastid origins in primary and secondary endosymbioses lineages. Engulfment of the cyanobacterium by a eukaryotic host resulted in the primary endosymbiosis lineages: glaucophytes, red algae, and green algae and land plants. Secondary endosymbiosis of a red alga by a eukaryotic cell resulted in five lineages with red algal-derived plastids. Similarly, green algae engulfed by eukaryotic cells resulted in euglenophytes and chlorarachniophytes, two lineages with green algal-derived plastids.

1.2.3 Plastid Evolutionary Origins

Primary plastids are believed to have arisen only once—within the common ancestor of green algae, red algae, and glaucophytes (i.e., the Archaeplastida) (Bhattacharya and Medlin 1995). Eukaryote-eukaryote derived plastids, however, have evolved multiple times independently, but the exact number is debated. There are many challenges associated with resolving the history and diversification of plastids, not the least of which is proving outright plastid loss (Green 2011), which likely occurred in multiple eukaryotic groups, including ciliates. Moreover, eukaryotic phylogenetic analyses are often hindered by "phylogenomic falsification"—variations in phylogenetic signals and the subsequent inconsistent (and sometimes erroneous) implications when using plastid rather than nuclear data, and vice versa (Burki *et al.* 2012). Noting that, in most cases, nuclear genomes are a fusion of genes that have not only been inherited vertically, but are also influenced by horizontal gene transfers (Green 2011). Lastly, there is an overall lack of genetic data from many key plastid-bearing lineages (Burki *et al.* 2012). All that said, our understanding of plastid evolution has improved immensely over the past decade and has led to a fundamental shift in how we view eukaryotic diversity (Smith and Lee 2014; Molina *et al.* 2014; Figueroa-Martinez *et al.* 2015).

1.3 The Loss of Photosynthesis

1.3.1 The general loss of photosynthesis

Acquiring a plastid and associated photosynthetic capabilities is no easy feat: it took more than a billion years of host-endosymbiont coevolution, integration, and regulation for the development of primary eukaryotic algae (Gould *et al.* 2008). Nevertheless, time and again, species have dispensed with photosynthesis, irrespective of how they acquired their plastid (Tartar and Boucias 2004; Gockel and Hachtel 2000). Throughout the eukaryotic tree there are numerous examples of organisms that have photosynthetic ancestors but that can no longer derive energy from sunlight. Even more intriguing: most species that lose photosynthesis still retain a nonphotosynthetic plastid and a plastid genome (Borza *et al.* 2005; Lim and McFadden 2010; Gould *et al.* 2008; Krause 2012). For example, only about half of plastid-bearing dinoflagellates are photosynthetic, and

although plastids are almost ubiquitous throughout the Apicomplexa, not a single species is photosynthetic (Gaines and Elbrachter 1987). There is also a number of primary-plastid-containing taxa that have lost photosynthesis but have retained a plastid along with a plastid genome (Table 1; DePamphilis and Palmer 1990; McNeal *et al.* 2007; DePamphilis *et al.* 1997; Delannoy *et al.* 2011; Wolfe *et al.* 1992), some of which are quite well known, such as certain orchid (*Rhizanthella gardneri*) and dodder (*Cuscuta exaltata*) species (DePamphilis and Palmer 1990; McNeal *et al.* 2007; Delannoy *et al.* 2011).

Table 1: Examples of nonphotosynthetic green algae and land plants including a description of their plastid genomes.

GREEN ALGAE					
Species Name	**Lifestyle**	**Known Host**	**Environment**	**Genomic Features**	**References**
***Helicosporidium* sp. (Chlorophyta: Trebouxiophyceae)**	Nonphotosynthetic, pathogenic, obligate parasitism, free-living cysts	Invertebrates (eg. insects)	Gut of host	37.5 kb, no photosynthetic genes, limited presence of non-coding DNA (1 intron and tiny intergenic spaces), no inverted repeats, no duplicated genes, highly structured	(de Koning and Keeling 2006; Tartar and Boucias 2004)
***Polytomella* spp. (Chlorophyta: Chlamydomondaceae)**	Nonphotosynthetic, free-living	N/A (not available)	Freshwater algae	N/A (possible loss of plastid genome)	(Smith and Lee 2014)
***Polytoma uvella* (Chlorophyta: Chlamydomondaceae)**	Nonphotosynthetic, free-living, saprophytic	N/A (not available)	Decaying matter, freshwater/ terrestrial algae	~75kb	(Figueroa-Martinez *et al.* 2015)
***Prototheca wickerhamii* (Chlorophyta: Trebouxiophyceae)**	Nonphotosynthetic, parasitic, free-living	Humans, animals, other vertebral hosts	Free-living, soil and aqueous habitats	No photosynthetic genes, sequences encode functions related to gene expression, reduced genome size (54.1 kb)	(Borza *et al.* 2005)
***Rhodochytrium spilanthidis* (Chlorophyta: Endosphaeraceae)**	Achlorophyllic, parasitic	Spilanthis, *Ambrosia artemisiaefolias* (ragweed)	Freshwater algae	N/A (not available)	(Atkinson 1908)

GREEN ALGAE					
Species Name	**Lifestyle**	**Known Host**	**Environ-ment**	**Genomic Features**	**References**
***Phyllosiphon arisari* (Chlorophyta: Phyllosiphonaceae)**	Chlorophyllic, endophytic, parasitic	Araceae (flowering plants)	Freshwater and marine waters	N/A	(Coute and Leclerc 1976; Aboal and Werner 2011)
***Cephaleuros parasiticus* (Chlorophyta: Trentopohliaceae)**	Parasitic, pathogenic, endophytic, no information on photosynthesis	Plants' stems, leaves, fruits, ex. tea plants, guava	Aerophilic, terrestrial	N/A	(Nelson 2008)
***Stomatochroon* sp. (Chlorophyta: Trentopohliaceae)**	Endophytic, has chloroplast, no information on photosynthesis	Plants, fruits, leaves	Terrestrial	N/A	(Brooks 2004)
LAND PLANTS (EMBRYOPHYTES)					
Species Name	**Lifestyle**	**Known Host**	**Environment**	**Genomic Features**	**References**
***Cryptothallus mirabilis* (AKA *Anerua mirabilis*, Aneuraceae, liverworts)**	Nonphotosynthetic, achlorophyllic, parasitic, obligate heterotrophic	Ectomycorrhizal association with fungi (Tulasnella)	Terrestrial	Complete loss of photosynthetic genes, highly reduced genome structure	(Wickett 2007)
***Rafflesia lagascae* (Rafflesiaceae, angiosperm)**	Nonphotosynthetic, parasitic	*Terastigma*	Terrestrial	N/A (possible loss of plastid genome)	(Molina *et al.* 2014)
***Epifagus virginiana* (Orobanchaceae, angiosperm)**	Nonphotosynthetic, parasitic	Beech trees	Terrestrial	Highly reduced, no photosynthetic genes	(Wolfe *et al.* 1992)

1.3.2 Loss of photosynthesis yet plastid retention hypotheses

Why would an organism discard such a valuable and hard-earned trait like photosynthesis? It is thought that many nonphotosynthetic species evolved from mixotrophic algae, which can survive heterotrophically and photoautrophically. Mixotrophy can be very advantageous because it allows an organism to live in a wide variety of environments, benefitting from both photosynthetic products and other sources of organic compounds. However, it is energetically taxing to maintain both the "heterotrophic" and "photoautotrophic" molecular machineries, and under certain conditions it might be more advantageous to dispose of one of the two machineries and funnel resources into a single feeding strategy, be it heterotrophism or photoautotrophism (Figueroa-Martinez *et al.* 2015). Some nonphotosynthetic algae—obligate heterotrophs—likely arose through such a scenario.

When species lose photosynthetic abilities, they almost always retain a plastid (Williams and Keeling 2003; Sekiguchi *et al.* 2002; DePamphilis and Palmer 1990; McNeal *et al.* 2007; Delannoy *et al.* 2011; Ralph *et al.* 2004). Indeed, plastids are the hub of many crucial cellular reactions apart from photosynthesis (Ralph *et al.* 2004; Smith *et al.* 2014; Borza *et al.* 2005; de Koning and Keeling 2004; Atteia *et al.* 2005). In plants and algae, for example, heme and fatty-acid biosynthesis, among many other biosynthetic and metabolic pathways, are reliant on plastid-located pathways.

Most of our understanding of nonphotosynthetic plastids and their functions comes from parasitic species, such as the malaria parasite *Plasmodium falciparum*, the trebouxiophyte green algae *Prototheca wickerhamii* (the causative agent of protothecosis) and *Helicosporidium* sp. (an insect gut parasite). In *P. wickerhamii*, for instance, which is closely related to *Chlorella* green algae, the plastid performs necessary metabolic functions (Borza *et al.* 2005). The 54.1 kilobase (kb) circular-mapping plastid genome of *P. wickerhamii* encodes mostly gene expression-related proteins (Knauf and Hachtel 2002). Enzymes responsible for plastid metabolic pathways such as starch biosynthesis, lipid metabolism, amino acid biosynthesis, purine biosynthesis, isoprenoid and tetrapyrrole biosynthesis, as well as a variety of oxidoreductive processes are nuclear-

encoded in nonphotosynthetic algae, including *P. falciparum*, are translated on cytosolic ribosomes, and posttranslationally imported into the plastid via various translocators and importers/exporters (Borza *et al.* 2005; Gould *et al.* 2008; Soll and Schleiff 2004; Hormann *et al.* 2007). In fact, more than 90% of these plastid metabolic enzymes are encoded within the nuclear genome and imported for plastid functions (Keegstra and Cline 1999). To put it in perspective, approximately 545 proteins are predicted to be apicoplast-targeted in *P. falciparum* (described further below). These are enzymes that are responsible for various metabolic pathways such as carbohydrate assimilation, lipid biosynthesis, tetrapyrrole and isoprenoids, and carotenoid biosynthesis. On the other hand, only 23 proteins are encoded by the apicoplast genome, or ~ 0.04% of enzymes and proteins that function within the apicoplast (Ralph *et al.* 2004). This is not a phenomenon that is unique to nonphotosynthetic organisms—it is also a property of photosynthetic organisms (albeit to a lesser degree), and is one of the defining features of endosymbiont-derived organelles.

1.4 Metabolic Pathways in Nonphotosynthetic Species

Metabolic pathways within the plastids of parasitic, nonphotosynthetic algae and plants have been highly characterized in recent years. Various hypotheses have linked plastid metabolic complexity to levels of parasitism and divergence from photosynthetic ancestry (de Koning and Keeling 2004). Obligate parasites are expected to possess less host autonomy (and thus can synthesize fewer metabolic pathways *de novo*) than those with free-living stages or cysts. Pombert *et al.* (2014) recently corroborated this by showing that gene families within the nuclear genome have been reduced within the nonphotosynthetic parasite *Helicosporidium* sp., but the alga has still retained nuclear gene families that ensure its host independence.

P. falciparum, commonly known as the malaria-causing parasite, possesses a plastid (also known as an apicoplast in apicomplexans) of red-algal secondary-endosymbiotic origins. Ralph *et al.* (2004) characterized some of its functions in hopes of finding a therapeutic drug that could target some of the proteins and enzymes that are found within the apicoplast, which are not normally found within their infected eukaryotic hosts. The study culminated in the classification of more than 500 proteins involved in housekeeping

activities, carbohydrate and starch biosynthesis, fatty acid biosynthesis, import machineries, and heme, tetrapyrrole, and isoprenoid biosynthesis (Table 2). One interesting finding, among many, involves the first steps of heme biosynthesis, which usually occurs in the plastids of plants and algae. However, *P. falciparum* uses the Shemin pathway by which glycine and succinyl-CoA are converted (via ALA synthase) to 5-aminolevulinic acid (or ALA) (Wilson *et al.* 1996; Varadharajan *et al.* 2002). Amino acid biosynthesis has not been shown to occur within the apicoplast, and the parasite may thus depend on the host for its amino acids to function (Ralph *et al.* 2004).

In *Helicosporidium*, an obligate parasite with a saprobic free-living life stage, the plastid is of green algal origins, and just like the *Plasmodium* apicoplast, is responsible for a variety of metabolic functions (Boucias *et al.* 2001; de Koning and Keeling 2004; Pombert *et al.* 2014). Fatty acid biosynthesis, tetrapyrrole and isoprenoids biosynthesis, reductive processes, and amino acid biosynthesis were found to occur within the parasite's plastid (Table 2; de Koning and Keeling 2004). Enzymes involved in fatty acid biosynthesis are targeted to the mitochondria as well as the plastid, similar to higher plants and green algae, such as *C. reinhardtii*. Similarly, *Helicosporidium* uses the C5 (or glutamate) pathway, relying on its plastid for the entirety of heme biosynthesis. Isoprenoid biosynthesis occurs via the DOXP pathway within the plastid, but with no retention of carotenoids (de Koning and Keeling 2004; Pombert *et al.* 2014). Three amino acids are predicted to be synthesized within the *Helicosporidium* plastid, namely serine, lysine, and leucine (de Koning and Keeling 2004). Recently, Pombert *et al.* (2014) also showed the presence of enzymes involved in the synthesis of phenylalanine, tyrosine, and tryptophan, produced via the shikimate pathway.

Perhaps the most advanced of these parasitic plastids is the one found within the trebouxiophyte green alga *P. wickerhamii*. The plastid is responsible for a wide range of metabolic pathways, including the ones described above for *Helicosporidium* and *P. falciparum* (with the exception of carotenoid biosynthesis), as well as pathways for approximately 10 amino acids. Synthesis of aromatic amino acids and branched amino acids as well as lysine, serine, threonine, and methionine are expected to occur in the plastids of *P. wickerhamii*, given the presence of various transcripts encoding these

metabolic enzymes with plastid-targeting signals (Table 2; Borza *et al.* 2005). Finally, *P. wickerhamii* keeps true to its green algal origins and synthesizes heme using the C5 pathway in the plastid, converting glutamate to 5-aminolevulinic acid—the first precursor of isoprenoid biosynthesis.

Table 2: Plastid metabolic pathways within nonphotosynthetic algae of primary or secondary endosymbiotic plastid origins. The presence of the pathway is indicated by ✓ whereas ✗ denotes the absence of the pathway.

Species	Starch	Fatty Acids	Amino Acids	Tetrapyrrole	Carotenoids
Plasmodium falciparum	✓	✓	✗	✓	✓
Helicosporidium	✓	✓	Limited (~6)	✓	✗
Prototheca wickerhamii	✓	✓	Limited (~10)	✓	✗

1.5 *Polytomella*

The *Polytomella* genus is composed of unicellular nonphotosynthetic freshwater green algae, and is closely related to the highly studied model organism *Chlamydomonas reinhardtii*. *Polytomella* is composed of four known lineages, represented by *P. parva*, *P. magna*, *P. piriformis*, and *P. capuana* (Figure 2). These plastid-containing species are found in Europe, with species isolated from an elm tree and a freshwater pond in England, Germany, and Italy, respectively (Pringsheim 1955; Moore *et al.* 1970; Brown *et al.* 1976). These nonphotosynthetic chlamydomonadaleans possess nuclear and mitochondrial genomes, the latter of which has been sequenced in all four lineages, but have lost their plastid genomes (Smith and Lee 2014).

Polytomella species are remarkable with respect to their mitochondrial genomes. They possess one of the smallest mitochondrial genomes in the Archaeplastida (13–28 kb, no introns, and 10 genes) (Smith *et al.* 2010). Their mitochondrial DNA (mtDNA) is linear and sometimes fragmented into multiple chromosomes, which contain telomeres (Smith *et al.* 2010). They are some of the richest species in GC content, with *P. capuana* being the highest (~57% GC content of mtDNA). Their mitochondrial genome exhibits high

substitution and mutation rates (leading to greater sequence divergence rates) in comparison with other chlamydomonadalean lineages (Smith *et al.* 2010).

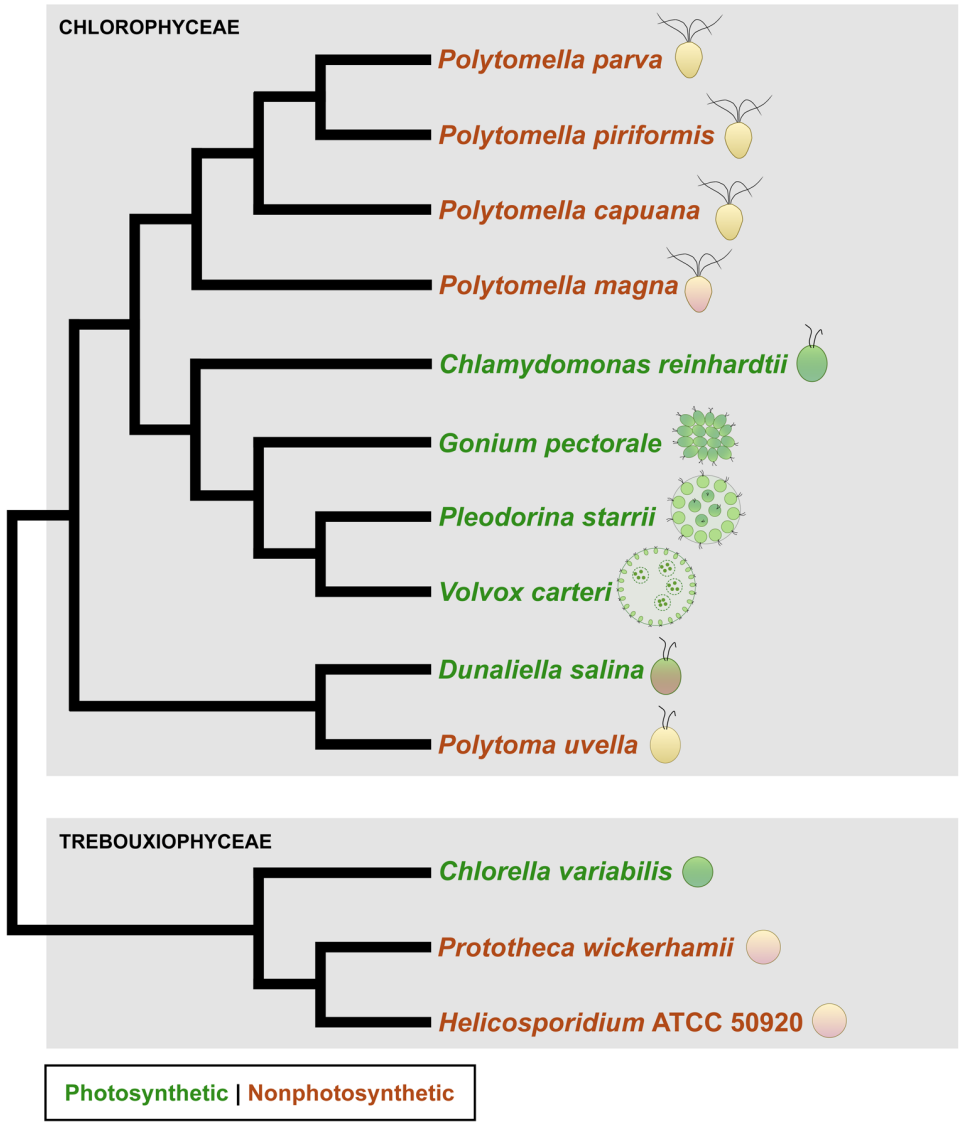


Figure 2: Tree of photosynthetic and nonphotosynthetic green algae, including *Polytomella* species. Branching order based on Smith and Lee (2014) and references therein.

One study has been published to highlight the plastid's function within these free-living green algae. Atteia *et al.* (2005) characterized the heme biosynthetic pathway within

Polytomella. 5'-aminolevulinic acid (or ALA), the precursor of heme, was shown to be synthesized through the glutamate pathway (C5 pathway) rather than the Shemin (C4) pathway suggesting that *Polytomella* spp. are dependent on the presence of the plastid for heme biosynthesis. As for the presence of a plastid genome in *Polytomella* spp., a recent study by Smith and Lee (2014) showed that these species have lost their plastid genomes altogether. Upon genome and transcriptome examination of the four *Polytomella* species, researchers failed to find evidence of plastid-encoded genes responsible for its plastid genome maintenance, replication, or translation/transcription. The paper suggests the absence of plastid-encoded genes, which was further confirmed by the absence of gene expression via transcriptome analysis. Smith and Lee (2014) did find a variety of nuclear-encoded plastid-targeted proteins responsible for a number of metabolic functions, such as heme biosynthesis, carbohydrate and starch metabolism, and fatty acid biosynthesis.

This instance of plastid genome loss is the first of its kind, in that no algal species with primary plastids (or any other type of plastid for that matter) have ever been shown to dispose of their organellar genome (Appendix A; Smith and Asmail 2014; Smith and Lee 2014). Some suggest that plastid genome loss might be inevitable, due to the highly oxidative environment of the plastid and its mutative effect on ptDNA (Wright *et al.* 2009). Others propose import machineries targeting nuclear tRNA glutamine ($tRNA^{Glu}$) into the plastid and beginning the process of heme biosynthesis, thus eliminating the necessity of a plastid DNA (ptDNA) to conduct this metabolic pathway (Smith and Lee 2014).

1.6 Thesis Rationale and Hypothesis

The goal of this thesis is to characterize the metabolic pathways within the plastids of the four known *Polytomella* species. I hoped to gain further insights into nonphotosynthetic species, specifically into the plastid metabolic pathways that persist versus those that are discarded, and whether the free-living *Polytomella* genus mirrors nonphotosynthetic species, or is more similar to photosynthetic green algae. Furthermore, the fact that *Polytomella* is closely related to *C. reinhardtii*, and the extensive characterization of the latter's enzymatic reactions and metabolic pathways, made my comparative genomics analyses easier.

The study is not only important in terms of its repertoire of the plastid metabolic pathways, but the study also reflects on evolutionary concepts and hypotheses regarding the loss of photosynthesis reported in other studies, validating some while refuting others. In doing so, it establishes *Polytomella* as an important organism for the study of genome-lacking plastid complexity and function in free-living, nonphotosynthetic green algae.

Hypothesis—

The plastids of the nonphotosynthetic *Polytomella* species carry out a diversity of metabolic functions similar to those of other closely related photosynthetic green algae.

Predictions—

1. *Polytomella* transcriptome and genome data analyses will reveal transcripts for biosynthetic pathways such as starch, heme and tetrapyrrole, and a variety of amino acids, and carotenoids.
2. *Polytomella* will not have any transcripts of genes encoded within the plastid genome.
3. *Polytomella* will not possess any photosynthesis-related transcripts.

1.7 References

Aboal, M., and Werner, O. 2011. Morphology, fine structure, lifecycle, and phylogenetic analysis of *Phyllosiphon arisari*, a siphonous parasitic green alga. Eur J Phycol 46: 181-192.

Adl, S.M., Simpson, A.G.B., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., Bowser, S.S. et al. 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol 52: 399-451.

Anderson, D.M. 1994. Red Tides. Sci Am 271: 52-58.

Archibald, J.M. 2007. Nucleomorph genomes: structure, function, origin, and evolution. Bioessays 29: 392-402.

Archibald, J.M. 2009. The puzzle of plastid evolution. Curr Biol 19: R81-R88.

Atkinson, G.F. 1908. A parasitic alga, *Rhodochytrium spilanthidis* Lagerheim in North America. Bot Gaz 46: 266-301.

Atteia, A., van Lis, R., and Beale, S.I. 2005. Enzymes of the heme biosynthetic pathway in the nonphotosynthetic alga *Polytomella* sp. Eukaryot Cell 4: 2087-2097.

Bendall, D.S. 2008. Introduction. Photosynthetic and atmospheric evolution. Philos Trans R Soc B 363: 2625-2628.

Bhattacharya, D., and Medlin, L. 1995. The phylogeny of plastids: a review based on comparisons of small-subunit ribosomal RNA coding regions. J Phycol 31: 489-498.

Boucias, D.G., Becnel, J.J., White, S.E., and Bott, M. 2001. In vivo and in vitro development of the protist *Helicosporidium* sp. J Eukaryot Microbiol 48: 460-470.

Bremer, K. 1985. Summary of green plant phylogeny and classification. Cladistics 1: 369-385.

Brooks, F.E. 2004. Plant-parasitic Algae (Chlorophyta: Trentopohliales) in American Samoa. Pacific Science 58: 419-428.

Brown, D.L., Massalski, A., and Patenaude, R. 1976. Organization of the flagellar apparatus and associate cytoplasmic microtubules in the quadriflagellate alga *Polytomella agilis*. J Cell Biol 69: 106-125.

Bullis, K. 2013. A material that could make solar power “dirt-cheap”. Energy News—MIT Technology Review.

Correa, J.A., and McLachlan, J.L. 1992. Endophytic algae of Chondrus crispus (rhodophyta). IV. Effects on the host following infections by *Acrochaete operculata* and *A. heteroclada* (Chlorophyta). Mar Ecol Prog Ser 81: 73-87.

Coute, A., and Leclerc, J.C. 1976. Revision of the systematic position of the parasitic algae *Phyllosiphon arisari* Kuehn after spectrophotometric analysis of its chlorophylls (Translated from: Revision de la position systematique de l'algue parasite Phyllosiphon arisari Kuehn d'apres l'analyse spectrophotometrique de ses chlorophylles). AGRIS from: http://agris.fao.org/agris-search/search.do?recordID=FR19770137193.

de Koning, A.P., and Keeling, P.J. 2004. Nucleus-encoded genes for plastid targeted proteins in Helicosporidium: functional diversity of a cryptic plastid in a parasitic alga. Eukaryot Cell 3: 1198-1205.

Douglas, S.E., and Penny, S.L. 1999. The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved synteny groups confirm its common ancestry with red algae. J Mol Evol 48:236-244.

Douglas, S.E., Zauner, S., Fraunhoza, M., Beaton, M., Penny, S., Deng, L., Wu, X., et al. 2001. The highly reduced genome of an enslaved algal nucleus. Nature 410:1091-1096.

Falkowski, P.G. 1994. The role of phytoplankton photosynthesis in global biogeochemical cycles. Photosynth Res 39: 235-258.

Figueroa-Martinez, F., Nedelcu, A.M., Smith, D.R., and Reyes-Prieto, A. 2015. When the lights go out: the evolutionary fate of free-living colorless green algae. New Phytol 206: 972-982.

Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkwoski, P. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. Science 281: 237-240.

Fuller, N.J., Campbell, C., Allen, D.J., Pitt, F.D., Zwirglmaier, K., Le Gall, F., Vaulot, D., and Scanlan, D.J. 2006. Analysis of photosynthetic picoeukaryote diversity at open ocean sites in the Arabian Sea using PCR biased towards marine algal plastids. Aquat Microb Ecol 43: 79-93.

Funes, S., Davidson, E., Reyes-Prieto, A., Magallón, S., Herion, P., King, M.P., and Gonzalez-Halphen, D. 2002. A green algal apicoplast ancestor. Science 298:2155.

Gaines, G., and Elbrachter, M. 1987. Heterotrophic nutrition. In F.J.R. Taylor [ed.], The biology of dinoflagellates, 224-267. Blackwell, Oxford, UK.

Gilson, P.R., Su, V., Slamovits, C.H., Reith, M.E., Keeling, P.J., and McFadden, G.I. 2006. Complete nucleotide sequence of the chloroarachniophyte nucleomorph: nature's smallest nucleus. Proc Natl Acad Sci USA 103: 9566-9571.

Gould, S.B., Waller, S.R., and McFadden, G.I. 2008. Plastid evolution. Annu Rev Plant Biol 59: 491-517.

Hallick, R.B., Hong, L., Drager, R.G., Favreau, M.R., Monfort, A., Orsat, B., Spielmann, A., and Stutz, E. 1993. Complete sequence of *Euglena gracilis* chloroplast DNA. Nucleic Acids Res 21:3537-3544.

Hansen, G., Botes, L., and De Salas, M. 2007. Ultrastructure and large subunit rDNA sequence of Lepidodinium viride reveal a close relationship to *Lepidodinium chlorophorum* comb. nov. (=*Gymnodiuium chlorophorum*). Phycol Res 55: 25-41.

Hedges, S.B., Blair, J.E., Venturi, M.I., and Shoe, J.L. 2004. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evol Biol 4:2.

Herrmann, R.G. 1997. Eukaryotism, towards a new interpretation. In HEA Schenk, RG Herrmann, KW Jeon, W Schwemmler, eds. Eukaryotism and Symbiosis, Springer, Heidelberg, Germany pp. 73-118.

Hibberd, D.J., and Norris, R.E. 1984. Cytology and ultrastructure of *Chloroarachnion reptans* (Chloroarachniophyta division nova, Chloroarachniophyceae classis nova). J Phycol 20:310-330.

Hoefnagels, M. 2011. Biology: concepts and investigations. McGraw-Hill. 2nd edition.

Hormann, F., Soll, J., and Bolter, B. 2007. The chloroplast protein import machinery: a review. Meth Mol Biol 390: 179-193.

Janouškovec, J., Horák, A., Oborník, M., Lukeš, J., and Keeling, P.J. 2010. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. Proc Natl Acad Sci USA 107: 10949-10954.

Keeling, P.J. 2010. The endosymbiotic origin, diversification and fate of plastids. Philos Trans R Soc Lond B Biol Sci 365: 729-748.

Keegstra, K., and Cline, K. 1999. Protein import and routing systems of chloroplasts. Plant Cell 11: 557-570.

Kim, E., Harrison, J., Sudek, S., Jones, M. D. M., Wilcox, H. M., Richards, T. A., Worden, A. Z., & Archibald, J. M. 2011. Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life. Proc Natl Acad Sci USA 108: 1496-1500.

Knauf, U., and Hachtel, W. 2002. The genes encoding subunits of ATP synthase are conserved in the reduced plastid genome of the heterotrophic alga *Prototheca wickerhamii*. Mol Genet Genomics 267: 492-497.

Krause, K. 2012. Plastid genomes of parasitic plants: a trail of reductions and losses. In Bullerwell C, ed., Organelle Genetics, Springer, Berlin, 79-103.

Leander, B.S., Triemer, R.E., and Farmer, M.A. 2001. Character evolution in heterotrophic euglenids. Europ J Protistol 37: 337-356.

Lemieux, C., Otis, C., and Turmel, M. 2007. A clade uniting the green algae Mesostigma viride and Chlorokybus atmosphyticus represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. BMC Biol 5:2.

Lim, L., and McFadden, G.I. 2010. The evolution, metabolism and functions of the apicoplast. Philos Trans R Soc Lond B Biol Sci 365: 749-763.

Martin, W., and Herrmann, R.G. 1998. Gene transfer from organelles to the nucleus: how much, what happens, and why? Plant Physiol 118: 9-17.

Moore, J., Cantor, M.H., Sheeler, P., and Kahn, W. 1970. The ultrastructure of *Polytomella agilis*. J Protozool 17: 671-676.

Molina, J., Hazzouri, K.M., Nickrent, D., Geisler, M., Meyer, R.S., Pentony, M.M., Flowers, J.M., et al. 2014. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). Mol Biol Evol doi: 10.1093/molbev/msu051.

Nedelcu, A.M. 2001. Complex patterns of plastid 16S rRNA gene evolution in nonphotosynthetic green algae. J Mol Evol 53:670—679.

Nelson, S.C. 2008. Cephaleuros species, the plant-parasitic green algae. Plant Disease PD-43.

Picket-Heaps, J.D., and Marchant, H.J. 1972. The phylogeny of the green algae: a new proposal. Cytobios 6:255—264.

Pombert, J.F., Blouin, N.A., Lane, C., Boucias, D., and Keeling, P.J. 2014. A lack of parasitic reduction in the obligate parasitic green alga *Helicosporidium*. PLOS Genet 10:e1004335. doi:10.1371/journal.pgen.1004355.

Ralph, S.A., van Dooren, G.G., Waller, R.F., Crawford, M.J., Fraunholz, M.J., Foth, B.J., Tonkin, C.J., Roos, D.S., and McFadden, G.I. 2004. Tropical infectious diseases: metabolic maps and functions of the *Plasmodium falciparum* apicoplast. Nat Rev Microbiol 2:203-216.

Roach.J. 2004. Source of Half Earth's Oxygen Gets Little Credit. Available at: http://news.nationalgeographic.com/news/2004/06/0607_040607_phytoplankton.html (last accessed 27 November 2014).

Sakamoto, W., Miyagishima, S., and Jarvis, P. 2008. Chloroplast biogenesis: control of plastid development, protein import, division and inheritance. The Arabidopsis Book, doi: 10.1199/tab.0110.

Sanchez-Puerta, M.V., Bachvaroff, T.R., and Delwiche, C.F. 2007. Sorting wheat from chaff in multi-gene analyses of chlorophyll c-containing plastids. Mol Phylogenet Evol 44:885-897.

Savage, L. 2013. Artificial photosynthesis: saving solar energy for a rainy day. Optics and Photonics News.

Sekiguchi, H., Moriya, M., Nakayama, T., and Inouye, I. 2002. Vestigial chloroplasts in heterotrophic stramenopiles *Pteridomonas danica* and *Ciliophrys infusionum* (Dictyochophyceae). Protist 153: 157-167.

Smith, D.R., Hua, J., and Lee, R.W. 2010. Evolution of linear mitochondrial DNA in three known lineages of *Polytomella*. Curr Genet 56: 427-438.

Smith, D.R., and Lee, R.W. 2014. A plastid without a genome: evidence from the nonphotosynthetic green algal genus *Polytomella*. Plant Physiol 164: 1812-1819.

Snounou, G., Viriyakosol, S., Jarra, W., Thaithong, S., and Brown, K.N. 1993. Identification of the four human malaria parasite species in field sample by the polymerase chain reaction and detection of a high prevalence of mixed infections. Mol Biochem Parasitol 58:283-292.

Soll, J., and Schleiff, E. 2004. Protein import into chloroplasts. Nat Rev Mol Cell Biol 5: 198-208.

Tartar, A., and Boucias, D.G. 2004. The non-photosynthetic, pathogenic green alga *Helicospordium* sp. has retained a modified, functional plastid genome. FEMS Microbiol Lett 233: 153-157.

Varadharajan, S., Dhanasekaran, S., Bonday, Z.Q., Rangarajan, P.N., and Padmanaban, G. 2002. Involvement of 5-aminolevulinate synthase encoded by the parasite gene in *de novo* haem synthesis by *Plasmodium falciparum*. Biochem J 367: 321-327.

Waller, R.F., Keeling, P.J., van Dooren, G.G., and McFadden, G.I. 2003. Comment on "a green algal apicoplast ancestor". Science 301:49a.

Watanabe, M.M., Suda, S., Inouye, I., Sawaguchi, I., and Chihara, M. 1990. *Lepidodinium viride* gen et sp. nov. (Gumnodiniales, Dinophyta), a green dinoflagellate with a chlorophyll a and b containing endosymbiont. J Phycol 26: 741-751.

Wickett, N.J. 2007. Plastid genome evolution of the non-photosynthetic liverwort Aneura mirabilis (Malmb.) Wickett and Goffinet (Aneuraceae). University of Connecticut Doctoral Dissertation, Paper ID: AAI3265818, available from: http://digitalcommons.uconn.edu/dissertations/AAI3265818/.

Williams, B.A., and Keeling, P.J. 2003. Cryptic organelles in parasitic protists and fungi. Adv Parasitol 54: 9-68.

Wilson, C.M., Smith, A.B., and Baylon, R.V. 1996. Characterization of the 5-aminolevulinate synthase gene homologue in *P. falciparum*. Mol Biochem Parasitol 75: 271-276.

Wolfe, K.H., Morden, C.W., and Palmer, J.D. 1992. Function and evolution of a minimal plastid genome from a non-photosynthetic parasitic plant. Proc Natl Acad Sci USA 89: 10648-10652.

Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G., and Bhattacharya, D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol 21:809-818.

Yoon, H.S., Hackett, J.D., Pinto, G., and Bhattacharya, D. 2002. The single, ancient origin of chromist plastids. Proc Natl Acad Sci USA 99: 15507-15512.

Chapter 2

2 The Starch Metabolic Pathway

2.1 Introduction: starch metabolism at a glance.

Starch and glycogen, the two major forms of carbohydrate/carbon storage, are found in organisms with diverse lineages and ancestry (Flatt 1995). Whereas starch is found in species with plastids derived from primary and/or secondary-endosymbiotic origins, glycogen is found in animals as well as fungi, bacteria, Achaea, and various heterotrophic protists. Glycogen consists of 8 to 10% α-1,6 branches, making a hydrosoluble particle that may extend up to a theoretical maximum of 40 nm in diameter. Starch, on the other hand, is insoluble and has two distinct polysaccharide fractions—amylose and amylopectin. Amylose makes up 10 to 30% of starch and consists of α-1, 4-glycosidic bonds with few α-1, 6 branches, whereas amylopectin forms the backbone of larger and more branched starch molecules that consist of more α-1, 6 branches and makes up the majority (70-80%) of starch (Streb and Zeeman 2012). Starch composition depends on environmental factors; for instance, nutrient-starved cultures of *Chlamydomonas reinhardtii* have been shown to possess approximately 15-30% amylose starch composition, while unstarved cultures make less than 5% amylose starch.

2.2 Starch biosynthesis pathway in detail.

In green algae and land plants, starch biosynthesis occurs within the cytoplasm and chloroplast. I set out to determine whether the starch biosynthetic pathway enzymes were present or absent within the four species of the *Polytomella* genus, to give us an initial hint of whether this pathway is still functional or has been obliterated within these nonphotosynthetic algae. There are a total of 11 enzymes involved in the starch biosynthetic pathway (Figure 3). The first substrate is glucose-6-phosphate, which is produced through glucose phosphorylation on the 6^{th} carbon, catalyzed by the enzyme hexokinase. Glucose-6-phosphate is interconverted into glucose-1-phosphate, by plastidial phosphoglucomutase, and creates ADP-glucose upon the latter's interaction with ATP, catalyzed by the enzyme ADP-glucose phosphorylase. The α-1,4 chain is then

elongated and branched through the subsequent addition of glucose residues from ADP-glucose via the work of starch synthases and starch branching enzymes, respectively.

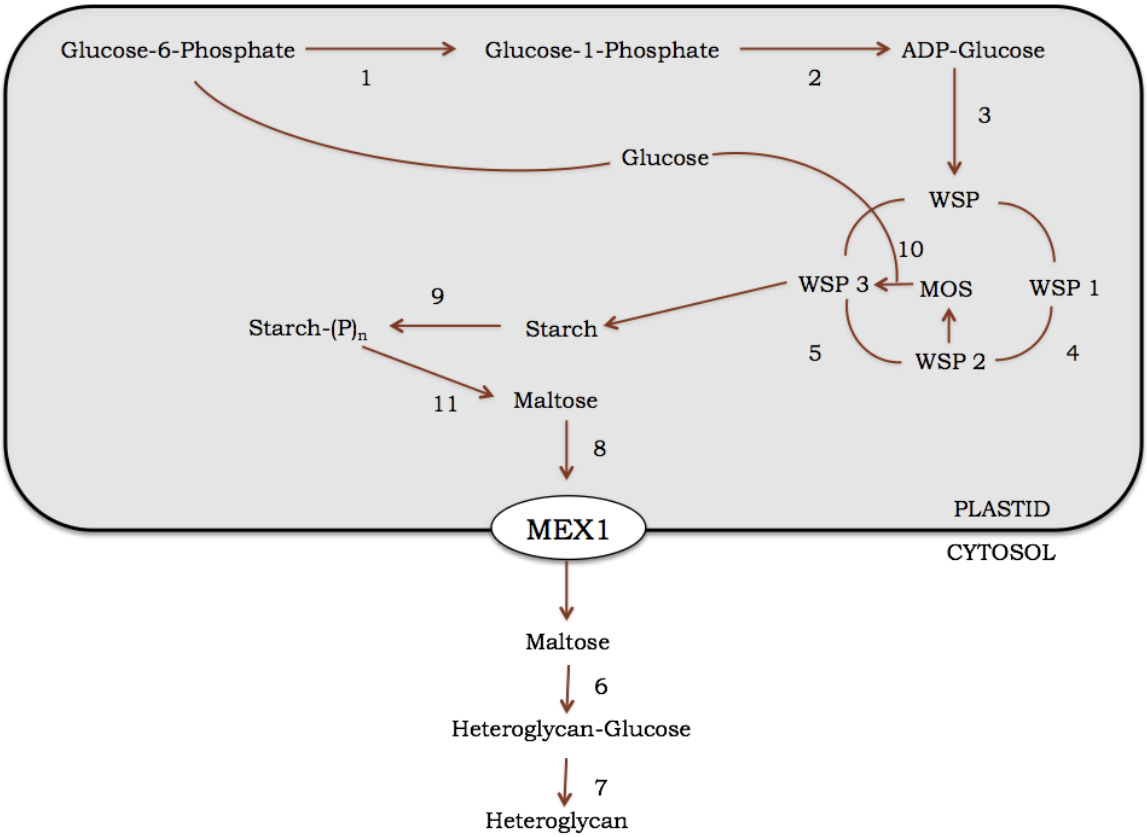


Figure 3: Consensus starch metabolic pathway in green algae. *WSP*, water soluble polysaccharides; *MOS*, malto-oligosaccharides; starch-$(P)_n$, polyphosphorylated starch. Numbers represent enzymes in order of their appearance within the pathway; 1, *GPM* or plastidial phosphoglucomutase; 2, *AGPase* or ADP-glucose pyrophosphorylase; 3, *SS* or starch synthase; 4, *BE* or branching enzyme; 5, *DBE* or debranching enzyme; 6, *GT* or α-1,4-glucanotransferase; 7, *SP* or starch phosphorylases; 8, *MEX1* or maltose-exporter like protein; 9, *GWD* or glucan-water dikinases; 10, *AMS* or α-amylases; 11, *BAMS* or β-amylases. Enzyme names and their corresponding GenBank accession number are listed in Section 2.6, and in the list of abbreviations.

Depending on whether amylose or amylopectin is being generated, the α-1,4 linked chain can be either branched, as mentioned above, or in fact, debranched via debranching

enzymes. The resulting starch is then phosphorylated, by glucan-water dikinases (GWD) (Figure 3). This is an essential step for starch degradation, as shown in leaves of land plant and other species (Lorberth *et al.* 1998). Mutants in GWD cause a very strong starch excess plastid phenotype in *Arabidopsis thaliana* (Caspar *et al.* 1991; Yu *et al.* 2001). The phosphorylated starch is then converted to maltose (a step catalyzed by β-amylases) to be transported into the cytoplasm, via the maltose exporter-like protein MEX1. Maltose can then be used by the cell, or converted into heteroglycan-glucose to be later converted into glucose for cellular expenditure or metabolite formation.

2.3 Methods: the transcriptomic and genomic data analyses of *Polytomella* spp.

2.3.1 Using *C. reinhardtii* as a reference organism

Chlamydomonas reinhardtii is a eukaryotic, unicellular photosynthetic green algal species that possesses two anterior flagella and one or more pyrenoids. It is a highly studied algal species due to its photosynthetic and heterotrophic properties. Moreover, the relative ease by which it and its mutants are cultured and isolated, the advanced molecular and genetic technologies developed, and the availability of such mutants, data, and protocols on public databases that can be found easily and free of charge (such as Phytozome v. 9, NCBI non-redundant database, and UNIPROT consortium) have made this organism a model organism (Goodstein *et al.* 2012; Pruitt *et al.* 2005; Apweiler *et al.* 2004).

C. reinhardtii possesses two traits that make it indispensible to this study: it is closely related to *Polytomella* and its nuclear genome and transcriptome have been fully sequenced (Merchant *et al.* 2007). Enzymes that function in the plastid to drive these metabolic pathways (starch, tetrapyrrole, amino acids, and carotenoid biosynthesis) are encoded within the nuclear genome and post-translationally imported into the plastid via various transporters and translocators (Gould *et al.* 2008; Soll and Schleiff 2004). What makes *C. reinhardtii* user-friendly is the wealth of information present and organized for each specific metabolic pathway, and the availability of such information on numerous public databases and scientific articles (Merchant *et al.* 2007; Harris and Stern 2009).

Transcripts encoding *C. reinhardtii* enzymes were downloaded as mRNA sequences from online databases, such as NCBI non-redundant database, using confirmed GenBank accession numbers (Harris and Stern 2009). These enzymes, which represent nuclear-encoded, plastid-targeted proteins, are expected to be conserved or highly similar in *Polytomella* due to their close phylogenetic relation to *C. reinhardtii* (Figure 2). In the case that a specific enzyme was not present within *C. reinhardtii*, mRNA sequences from other closely related organisms were downloaded instead. These species included *Volvox carteri* and *A. thaliana*.

2.3.2 Transcriptomic analyses

P. piriformis, P. parva, P. capuana, and *P. magna* (Göttingen Culture Collection of Algae, SAG, strains 63-10, 63-3, 63-5, and 63-9, respectively) were axenically grown and harvested as previously described (Smith and Lee 2014). *P. parva* RNA extraction, library preparation, Illumina sequencing, and *de novo* RNA-Seq assembly were carried out by the National Center for Genome Resources (NCGR) following the protocols of the MMETSP (Figure 4; Smith and Lee 2014; Keeling *et al.* 2015). For *P. piriformis*, *P. capuana* and *P. magna*, total cellular RNA was isolated using the Qiagen (MD, USA) RNeasy Plant Mini Kit and treated with Qiagen RNase-free DNase followed by RNA-Seq library preparation and Illumina (HiSeq 2500) sequencing (paired end, 2 x 150 cycle run) at the McGill University and Genome Quebec Innovation Centre (MUGQIC). The raw RNA-Seq data were trimmed and clipped with Trimmomatic (Bolger *et al.* 2014) and the normalized reads were assembled with Trinity (Haas *et al.* 2014) by the MUGQIC Bioinformatics team using their standard parameters.

Transcript and genome sequences for the four *Polytomella* species were made available for analyses via previous sequencing projects—described in MacDonald and Lee (2015). The DNA sequences were derived from a single cell per species that were cultured, crushed, and then had their total DNA extracted using known protocols. A similar procedure was done for the RNA extraction, although it should be noted that the same cells were not used for both processes. The DNA and RNA extracted were sequenced using Illumina HiSeq 2000 resulting in raw reads for both (Figure 4). RNA raw reads were trimmed and assembled into contigs, which were also available in annotated

versions. There are a total of 30 126 contigs for *P. parva*, 58 182 contigs for *P. magna*, 94 587 contigs for *P. capuana*, and 64 852 contigs for *P. piriformis*. These contigs ranged from a minimum of 150 nucleotides to a maximum of 14 173 nt (Table 3; Figure 4).

Table 3: *Polytomella* RNA-Seq data summary, also shown in Figures 4 and 5.

	P. piriformis	*P. parva*	*P. capuana*	*P. magna*
RNA-SEQ DATA[1]				
Number of reads	6.97×10^7	5.29×10^7	9.31×10^7	26.0×10^7
GenBank accession	SRX710732	SRX551283	SRX710731	SRX710730
Contigs	6.48×10^4	3.01×10^4	9.45×10^4	5.81×10^4
N50	1.28×10^4	5.99×10^4	2.45×10^4	1.72×10^4

Geneious v. 8.0.3 (Kearse *et al.* 2012) was used to construct local BLAST databases composed of the four *Polytomella* spp. RNA contigs. These databases were subjected to a search using the *C. reinhardtii* enzymes as query sequences, with homologous sequences representing hits within the *Polytomella* transcriptome. This was done using BLAST functions, specifically tBLASTx to look for *Polytomella* RNA contigs that were homologous to the mRNA sequences encoding the metabolic enzymes in *C. reinhardtii*. As depicted in Figure 5, an e-value of 1E-10 was used as the cut-off value, above which, the RNA contig was considered a poor fit and discarded from the dataset. Hits were double-checked against NCBI's non-redundant protein sequence database using BLASTx to confirm the function of the enzyme based on hits to other closely related species (Altschul *et al.* 1990). As such, hits represented the presence of the transcript for each evaluated metabolic pathway. No hits or irrelevant hits represent the absence of such a transcript, and therefore the absence of the corresponding enzyme in *Polytomella*. For instance, chlorophyllase involved in chlorophyll degradation, shows hits corresponding to cyclin-dependent kinase 20 or tRNA guanosine-2-O-methyltransferase. The latter is an enzyme involved in methionine to homocysteine interconversion, and has nothing to do with tetrapyrrole biosynthesis. Such an irrelevant hit is used to conclude that the specific enzyme chlorophyllase or its transcript, specifically, is missing from the *Polytomella* transcriptome.

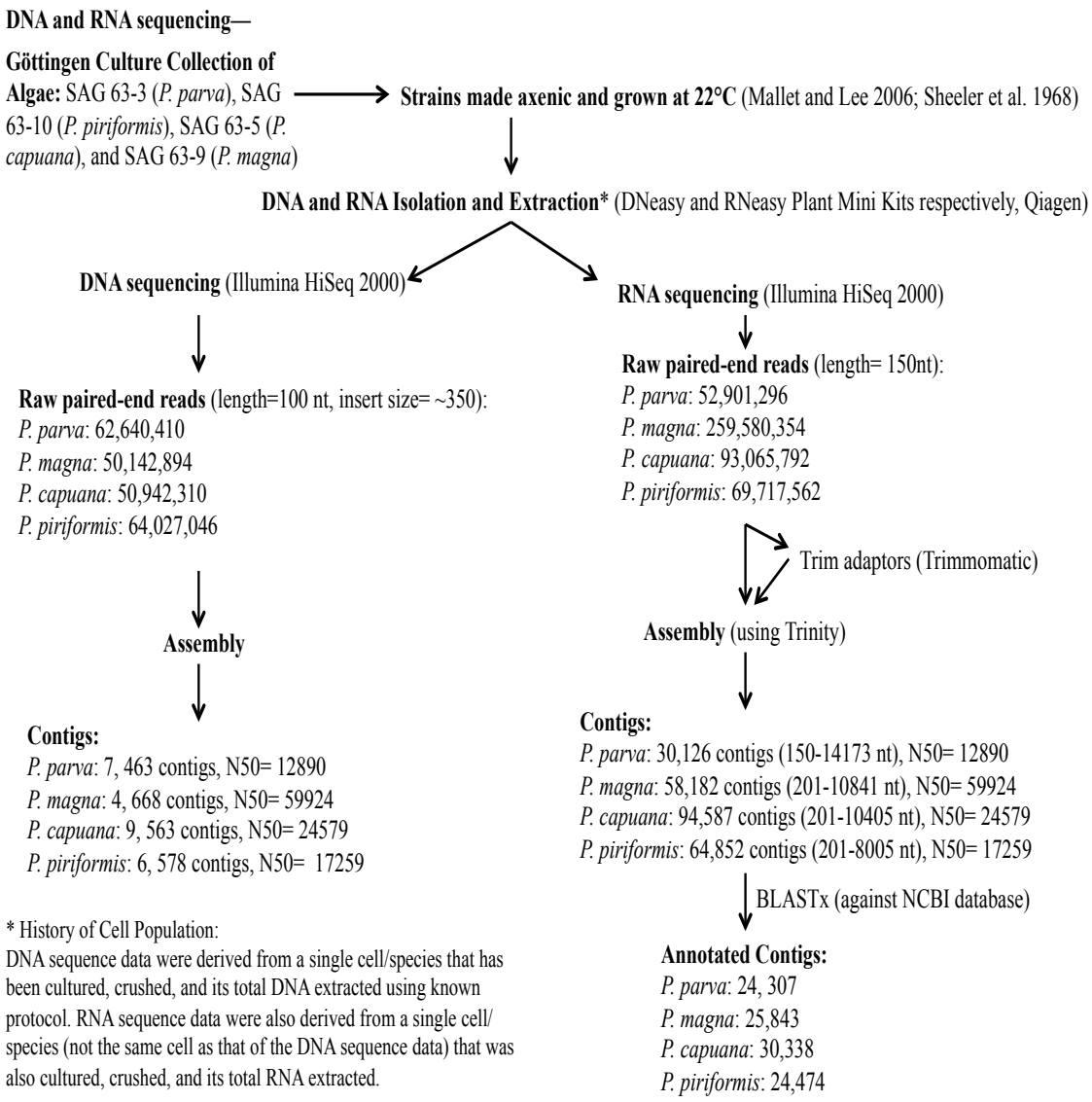
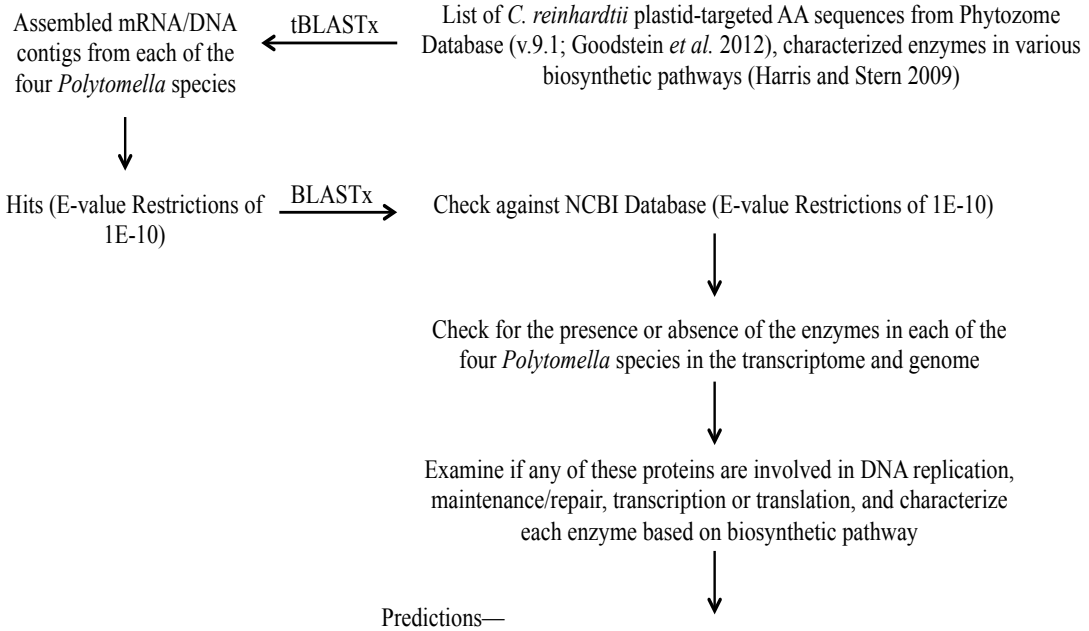


Figure 4: Outline of the DNA and RNA sequencing project for *Polytomella* spp. The four cultures were obtained from the Göttingen Culture Collection of Algae. The cells were subjected to established DNA and RNA extraction protocols. Illumina sequencing resulted in raw paired-end reads which were then assembled into contigs. RNA contigs were also annotated using BLASTx.



Predictions—

Functions	*C. reinhardtii*	*Polytomella* sp.
Photosynthesis	✓	×
Amino acid metabolism	✓	✓
Tetrapyrrole biosynthesis	✓	✓
DNA replication	✓	×
DNA transcription and RNA translation	✓	×
Carbon metabolism	✓	✓
Lipid Metabolism	✓	✓
Protein-related functions	✓	✓

Figure 5: Data analyses summary for the four *Polytomella* spp. Established pathway-specific enzymes in *C. reinhardtii* and other closely related species were used to search the *Polytomella* transcriptome and genome for homologous sequences, that would confer the same function/ enzyme. Hits were checked against NCBI databases for further confirmation. It is predicted that no genes involved in photosynthesis or plastid genome expression are present within *Polytomella* spp. In contrast, all other metabolic pathways are expected to occur within the *Polytomella* plastid, similar to other closely related green algae.

2.3.3 Genomic Analyses

In cases where there were no or irrelevant hits, the *Polytomella* genome (i.e., DNA) was checked to see if it contains genes for the various enzymes. For this, the same mRNA *C.*

reinhardtii enzyme sequences were used, but the local database consisted of the four *Polytomella* species DNA contigs in this case, rather than the RNA contigs used previously. Similarly, an e-value of 1E-10, matrix: BLOSUM62, maximum word size of 3, and a standard genetic code, were used as hit restrictions. Hits meant that the enzyme was encoded within the genome, while a no hit or irrelevant hits result meant that the enzyme was not encoded within the genome. For each, whether a hit or not, the contigs were rechecked with NCBI to confirm the function of the enzyme encoded by that gene in other closely related species that have previously characterized the enzyme.

2.4 Results

Genome and transcriptome analyses showed that all of the genes encoding enzymes involved in starch biosynthesis are present at both the transcriptomic and genomic levels in the four *Polytomella* species. The only missing component is the maltose-exporter like protein MEX1, responsible for maltose transport from the plastid compartment into the cytoplasm (Table 4). The gene was missing entirely from the transcriptome and genome data. Intriguingly, enzymes and proteins responsible for maltose conversion into heteroglycan-glucose, following its transport into the cytoplasm, are present at the genome and transcriptome levels.

Table 4: Starch metabolic pathway enzymes, shown in order of appearance within the pathway. The presence of the enzyme is indicated by ✓ whereas ✗ denotes the absence of the enzyme. Enzymes that were absent within the transcriptome were also checked at the genome level, in which case ✗/✗ denotes the absence of the enzyme at both transcriptome and genome levels.

Step	ENZYMES	*P. capuana*	*P. magna*	*P. parva*	*P. piriformis*
1	GPM	✓	✓	✓	✓
2	AGPase	✓	✓	✓	✓
3	SS	✓	✓	✓	✓
4	BE	✓	✓	✓	✓
5	DBE	✓	✓	✓	✓
6	GT	✓	✓	✓	✓
7	SP	✓	✓	✓	✓
8	MEX1	✗/✗	✗/✗	✗/✗	✗/✗
9	GWD	✓	✓	✓	✓

10	AMS	✓	✓	✓	✓
11	BAMS	✓	✓	✓	✓

2.5 Discussion

All of the starch biosynthetic enzymes are present, both at the transcript and the genome levels. In contrast, all four species were missing the MEX1 protein. Whereas starch is synthesized within the *Polytomella* plastid, it is degraded into glucose and maltose via glucan-water dikinases by phosphorylation. However, the absence of a MEX1 enzyme suggests that the starch degradation product, maltose, is unable to get to the cytoplasm (Figure 3). This contrasts with the fact that enzymes downstream of this transport step are present, within the *Polytomella* transcriptomes and genomes. Maltose may therefore, have other transporters that are able to transport it into the cytoplasm or it could be exported from the plastid as isomaltose, but such a claim would need to be investigated experimentally using HPLC techniques (Servaites and Geiger 2002; Weise *et al.* 2003). We set to determine if there are any other types of transporters present in the interplastid membrane and we found two other types of transporters. Two glucose-6-phosphate translocators (accession no. NM_124861 and AY081479, respectively) and one hexose/glucose transporter (accession no. AF215855) were found in all four *Polytomella* species at the transcriptome level. These are glucose transporters rather than maltose ones, however, and the two have been shown to possess separate and independent transporters (Herold *et al.* 1981; Rost *et al.* 1996). This may not represent all of the transporters/exporters/ translocators present within the plastid membrane, but conveys that there are more transporters that need to be considered. It is interesting to note that mutants with defective maltose and glucose transporters cause starch degradation products to accumulate and disrupt normal plastid function leading to achlorotic phenotypes, decreased starch content, and decreased photosynthetic rates (Cho *et al.* 2011). In some ways, that is the case with the *Polytomella* genus, which is nonphotosynthetic with an achlorotic plastid.

Nonetheless, MEX1 absence at the transcriptome and genome level may open up a biofuel implication to the study of *Polytomella*. MEX1 mutants in *A. thaliana* have been shown to accumulate more starch (~40 times more than WT) within their chloroplasts, as

well as synthesize more starch overall (Niittylä *et al.* 2004; Lu and Sharkey 2006; Radakovits *et al.* 2010). Such a result supports the accumulation of starch granules using light and transmission electron microscopy (TEM) imaging of the *Polytomella* plastid, in the cells' anterior portion, often masking the nucleus from view (Pringsheim 1955).

It was intriguing to determine where the glucose-6-phosphate was coming from into the plastid. Glucose-6-phosphate can come from three sources: glucose phosphorylation catalyzed by hexokinase, from glycogen breakdown or glucose-1-phosphate catalyzed by phosphoglucomutase, or from fructose-6-phosphate involved in the Calvin cycle or glycolysis catalyzed by phosphoglucose isomerase. All three enzymes were found within the *Polytomella* transcriptome. This suggests that glucose-6-phosphate either enters the plastid through the glucose-6-phosphate translocator mentioned above, is produced through glucose phosphorylation, or is produced via glycolysis. Any enhancement of these three steps can further encourage starch biosynthesis within the plastids of the *Polytomella* species.

It is more desirable to have a naturally occurring biofuel-favorable trait, rather than to specifically genetically engineer it. Moreover, *Polytomella* as any other unicellular microalgae provides several substantial benefits for biofuel use. These include their availability year long for harvest, smaller footprint, and adaptability to a wide range of environmental conditions, and their relative ease for genetic modification towards more efficient biofuel applications (Dismukes *et al.* 2008; Falkowski and Raven 1997; Long *et al.* 1994; Grossman *et al.* 2007). *C. reinhardtii* and *Volvox carteri*, two closely related species to *Polytomella*, are some of the few algae to be genetically modified with stable transcript expression and transgenic DNA integration, providing a point of reference for such studies and experiments in *Polytomella* (Zorin *et al.* 2009; Molnar *et al.* 2009; Zhao *et al.* 2009; Rosenberg *et al.* 2008; Beer *et al.* 2009). What is interesting is that in other heterotrophic species, such as *Chlorella protothecoides*, the rate of fatty acid and oil accumulation is 3-4 times more than phototrophic cells (Wu and Miao 2006), providing further implications as to whether it would be the case in *Polytomella* as well. It would be interesting to further study and experiment with nonphotosynthetic, free-living green

algae such as *Polytomella*, not only in terms of their starch metabolic pathway, but also to look at other metabolic pathways with mutants and experimental knockouts.

2.6 Abbreviations

1. GPM/ PGM: Plastidial phosphoglucomutase
2. AGPase: ADP-glucose pyrophosphorylase
3. SS: Starch synthases
4. BE: Branching enzyme
5. DBE: Debranching enzyme
6. GT: α-1, 4-glucanotransferases
7. SP: Starch phosphorylases
8. MEX1: Maltose exporter-like protein
9. GWD: Glucan-water dikinases
10. AMS: α-amylases
11. BAMS: β-amylases

2.7 References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. J Mol Biol 215: 403-410.

Apweiler, R., Bairoch, A., and Wu, C.H. 2004. Protein sequence database. Curr Opin Chem Biol 8: 76-80.

Beer, L.L., Boyd, E.S., Peters, J.W., and Posewitz, M.C. 2009. Engineering algae for biohydrogen and biofuel production. Curr Opin Biotechnol 20: 264-271.

Bolger, A.M., Lohse, M., Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics btu170.

Caspar, T., Lin, T.P., Kakefuda, G., Benbow, L., Preiss, J., and Somerville, C. 1991. Mutants of *Arabidopsis* with altered regulation of starch degradation. Plant Physiol 95: 1181-1188.

Cho, M.H., Lim, H., Shin, D.H., Jeon, J.S., Bhoo, S.H., Park, Y.I., and Hahn, T.R. 2011. Role of the plastidic glucose translocator in the export of starch degradation products from chloroplasts in *Arabidopsis thaliana*. New Phytol 190: 101-112.

Colleoni, C., Linka, M., Deschamps, P., Handford, M.G., Dupree, P., Weber, A.P.M., Ball, S.G. 2010. Phylogenetic and biochemical evidence supports the recruitment of an ADP-glucose translocator for the export of photosynthate during plastid endosymbiosis. Mol Biol Evol 27: 2691-2701.

Dismukes, G.C., Carrieri, D., Bennette, N., Ananyev, G.M., and Posewitz, M.C. 2008. Aquatic phototrophs: efficient alternatives to land-based crops for biofuels. Curr Opin Biotechnol 19: 235-240.

Falkowski, P., Raven, J. 1997. Aquatic photosynthesis. Malden: Blackwell.

Flatt, J.P. 1995. Use and storage of carbohydrate and fat. Am J Clin Nutr 61: 952S-959S.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., et al. 2012. Phytozome: a comparative platform for green plant genomics. Nucl Acids Res 40: D1178-D1186.

Gould, S.B., Waller, S.R., and McFadden, G.I. 2008. Plastid evolution. Annu Rev Plant Biol 59: 491-517.

Grossman, A.R., Croft, M., Gladyshev, V.N., Merchant, S.S., Posewitz, M.C., Prochnik, S., and Spalding, M.H. 2007. Novel metabolism in *Chlamydomonas* through the lens of genomics. Curr Opin Plant Biol 10: 190-198.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., *et al.* 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8: 1494-1512.

Harris, E.H., and Stern, D.B. 2009. The Chlamydomonas Sourcebook, Organellar and Metabolic Processes. 2nd ed. Academic Press, Canada.

Herold, A., Leegood, R.C., McNeil, P.H., and Robinson, S.P. 1981. Accumulation of maltose dueing photosynthesis in protoplasts isolated from spinach leaves treated with mannose. Plant Physiol 67:85-88.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647-1649.

Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., *et al.* 2014. The marine microbial eukaryote transcriptome sequencing project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol 12: e1001889.

Long, S.P., Humphries, S., Falkowski, P.G. 1994. Photoinhibition of photosynthesis in nature. Annu Rev Plant Physiol Plant Mol Biol 45:633-662.

Lorberth, R., Ritte, G., Willmitzer, L., and Kossmann, J. 1998. Inhibition of a starch-granule-bound protein leads to modified starch and repression of cold sweetening. Nature Biotechnol 16: 473-477.

Lu, Y., and Sharkey, T.D. 2006. The importance of maltose in transitory starch breakdown. Plant Cell Environ 29: 353-366.

MacDonald, S.M., Lee, R.W. 2015. Validation of *Polytomella piriformis* nomen nudum (Chlamydomonadaceae): a distinct lineage within a genus of Non-Photosynthetic Green Algae. J Eukaryot Microbiol doi: 10.1111/jeu.12241.

Mallet, M.A., and Lee, R.W. 2006. Identification of three distinct *Polytomella* lineages based on mitochondrial DNA features. J Eukaryot Microbiol 53: 79—84

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., et al. 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science 318: 245-251.

Molnar, A., Bassett, A., Thuenemann, E., Schwach, F., Karkare, S., Ossowski, S., Weigel, D., and Baulcombe, D. 2009. Highly specific gene silencing by artificial microRNAs in the unicellular alga *Chlamydomonas reinhardtii*. Plant J 58: 165-174.

Niittylä, T., Messerli, G., Trevisan, M., Chen, J., Smith, A.M., and Zeeman, S.C. 2004. A previously known maltose transporter essential for starch degradation in leaves. Science 303: 87-89.

Pringsheim, E.G. 1955. The Genus *Polytomella*. J Protozool 2: 137-145.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. 2005. NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 33: D501-D504.

Radakovits, R., Jinkerson, R.E., Darzins, A., and Posewitz, M.C. 2010. Genetic engineering of algae for enhanced biofuel production. Eukaryot Cell 9:486-501.

Rosenberg, J.N., Oyler, G.A., Wilkinson, L., and Betenbaugh, M.J. 2008. A green light for engineered algae: redirecting metabolism to fuel a biotechnology revolution. Curr Opin Biotechnol 19: 430-436.

Rost, S., Frank, C., and Beck, E. 1996. The chloroplast envelope is permeable for maltose but not for maltodextrins. Biochem Biophys Acta 1291: 221-227.

Servaites, J.C., and Geiger, D.R. 2002. Kinetic characteristics of chloroplast glucose transport. J Exp Bot 53: 1-11.

Sheeler, P., Cantor, M., and Moore, J. 1968. Growth characteristics of *Polytomella agilis* in batch culture. Life Sci 7: 289 -293.

Smith, D.R., and Lee, R.W. 2014. A plastid without a genome: evidence from the nonphotosynthetic green algal genus *Polytomella*. Plant Physiol 164: 1812-1819.

Soll, J., and Schleiff, E. 2004. Protein import into chloroplasts. Nat Rev Mol Cell Biol 5: 198-208.

Streb, S., and Zeeman, S.C. 2012. Starch metabolism in *Arabidopsis*. The Arabidopsis Book.

Weber, A.P.M. 2004. Solute transporters as connecting elements between cytosol and plastid stroma. Curr Opin Plant Biol 7: 247–253.

Weber, A., Servaites, J.C., Geiger, D.R., Kofler, H., Hille, D., Gröner, F., Hebbeker, U., Flu¨gge, U.I. 2000. Identification, purification, and molecular cloning of a putative plastidic glucose translocator. Plant Cell 12: 787–802.

Weise, S.E., Weber, A.P.M., and Sharkey, T.D. 2004. Maltose is the major form of carbon exported from chloroplast at night. Planta 218:474-482.

Wu, Q., and Miao, X. 2006. Biodiesel production from heterotrophic microalgal oil. Bioresour Technol 97: 841-846.

Yu, T.S., Kofler, H., Häusler, R.E., Hille, D., Flügge, U.I., Zeeman, S.C., Smith, A.M., Kossman, J., Lloyd, J., et al. 2001. The *Arabidopsis* sex1 mutant is defective in the R1 protein, a general regulator of starch degradation in plants, and not in the chloroplast hexose transporter. Plant Cell 13: 1907-1918.

Zhao, T., Wang, W., Bai, X., and Qi, Y. 2009. Gene silencing by artificial microRNAs in *Chlamydomonas*. Plant J 58: 157-164.

Zorin, B., Lu, Y., Sizova, I., and Hegemann, P. 2009. Nuclear gene targeting in *Chlamydomonas* as exemplified by disruption of the PHOT gene. Gene 432: 91-96.

Chapter 3

3 Tetrapyrrole Biosynthetic Pathway

3.1 Introduction: tetrapyrrole biosynthesis at a glance.

The tetrapyrrole biosynthetic pathway is involved in the production of light-harvesting pigments, such as chlorophylls and carotenoids, as well as prosthetic groups of respiratory proteins, including hemes and sirohemes (Ralph *et al.* 2004; de Koning and Keeling 2004). The pathway is deemed crucial given its widespread retention in photosynthetic and nonphotosynthetic species alike. Whereas all tetrapyrroles share a common 5-aminolevulinic acid (ALA) precursor, the pathway can start via two routes, depending on whether the organism in question is a photosynthetic, or a free-living/parasitic plastid-bearing nonphotosynthetic species (Papenbrock and Grimm 2001).

In plants, green algae, and bacteria, ALA is formed from glutamine or L-glutamic acid (Glu) brought into the plastid via a $tRNA^{Glu}$ in a three-step reaction (Kannangara *et al.* 1988; Papenbrock and Grimm 2001). Glutamine is ligated onto $tRNA^{Glu}$, enzymatically activating the glutamine, followed by the reduction of the α-carboxyl group of the activated glutamine into an aldehyde forming Glu-1-semialdehyde (or GSA), and lastly, the transamination of GSA to form its isomer ALA (Kannangara *et al.* 1984; Pontoppidan and Kannangara 1994; Kannangara and Gough 1978). Three enzymes are responsible for the steps mentioned above, namely: Glu-tRNA synthetase (GTS), Glu-tRNA reductase (GTR), and Glu-1-semialdehyde aminotransferase (GSAT), respectively.

In contrast, nonphotosynthetic eukaryotes synthesize ALA using mitochondrial enzymes in a condensation reaction of succinyl-CoA with glycine via the Shemin pathway (Gibson *et al.* 1958; Kikuchi *et al.* 1958; Papenbrock and Grimm 2001). The reaction is catalyzed by the enzyme δ-aminolevulinate synthase (ALAS), which is of α-proteobacterial origins and is found in a wide range of eukaryotes, with the exception of species with primary plastids. However, various other plastid-bearing eukaryotes do contain ALAS, the product of which (ALA) is transferred into the plastid for further synthetic reactions

leading to heme (Ralph *et al.* 2004; Sato *et al.* 2004; Varadharajan *et al.* 2002; Panek and O'Brian 2002).

3.2 Tetrapyrrole biosynthetic pathway in detail.

The tetrapyrrole biosynthetic pathway for green algae is summarized in Figure 6. There are a total of 39 enzymes involved in this pathway, which make up 15 steps, starting from glutamate and ending up with chlorophylls (a) and (b). All chlorophylls are made up of two components: the tetrapyrrole pigment, whose carbon atoms are derived from glutamate within the chloroplast/plastid, and an isoprenoid phytol. These components covalently join within the later stages of tetrapyrrole biosynthesis, forming the major chlorophylls (a) and (b) (Beale 2009).

Generally, the first six steps result in the generation of the tetrapyrrole uroporphyrinogen III, and the further steps lead to either a second tetrapyrrole called protoporphyrin IX or result in sirohemes, which represents the first branch point of the tetrapyrrole biosynthetic pathway (Figure 6). In the case that uroporphyrinogen III leads to protoporphyrin IX, the latter could either lead to chlorophyll formation further downstream in the pathway, or could branch out to form heme. Insertion of magnesium into protoporphyrin IX leads to the formation of an isocyclic ring that is present in all chlorophylls. Then, chlorin formation and vinyl to ethyl reduction result in chlorophyllide a, which (when joined with phytol) results in chlorophyll (a). Chlorophyll (b) is derived from the same precursor, chlorophyllide (a), which is converted to chlorophyllide (b), resulting in the production of chlorophyll (b) (Figure 6).

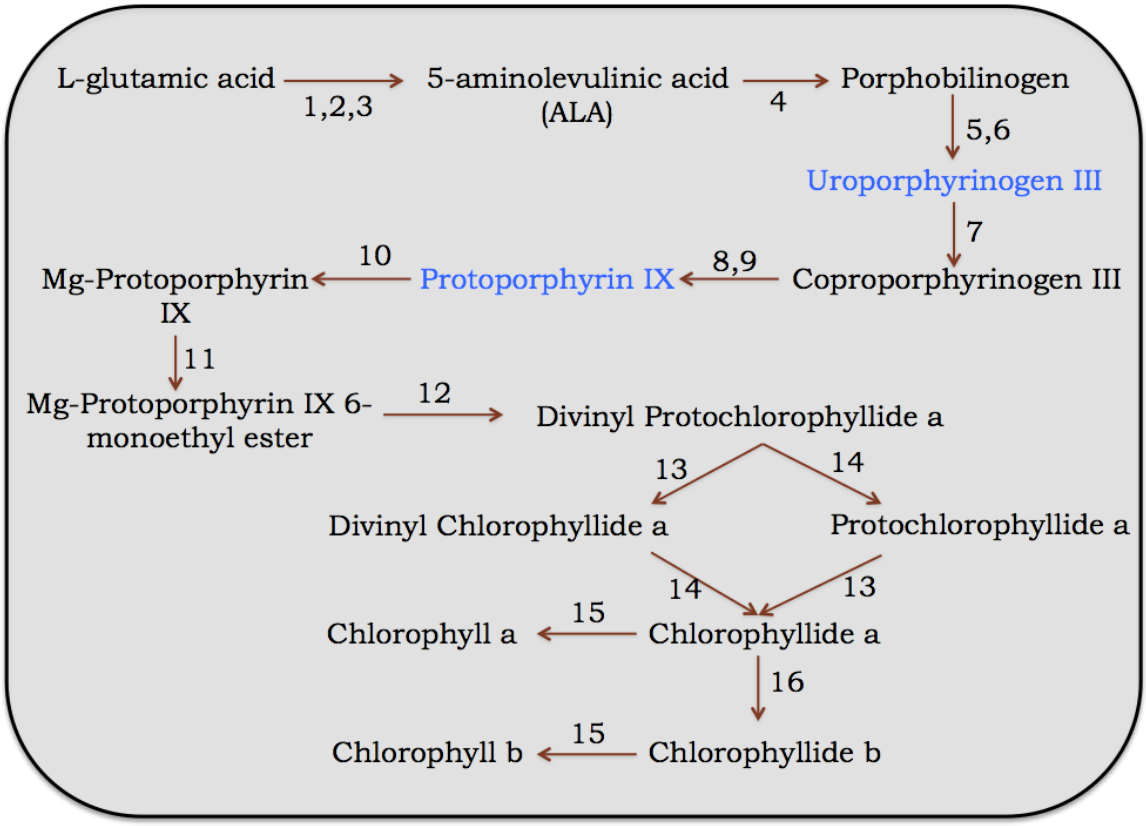


Figure 6: Tetrapyrrole biosynthetic pathway. Blue text refers to branch points leading to further siroheme and heme biosynthesis in the mitochondria. Numbers represent catalytic steps with their corresponding enzymes; 1, *GTS* or glutamyl-tRNA synthetase; 2, *GTR* or glutamyl-tRNA reductase; 3, *GSAT* or glutamate-1-semialdehyde aminotransferease; 4, *ALAD* or porphobilinogen synthase (ALA dehydratase); 5, *PBGD* or hydroxymethylbilane synthase; 6, *UROS* or uroporphyrinogen III synthase; 7, *UROD* or uroporphyrinogen III decarboxylase; 8, *CPX* or coproporphyrinogen III oxidative decarboxylase; 9, *PPX* or protoporphyrinogen IX oxidase; 10, *CHLD*/ *CHLH*/ *CHLI*/ *GUN4* or Mg-chelatase and Mg-chelatase activating protein; 11, *PPMT* or Mg-protoporphyrin IX methyltransferase; 12, *CRD*/ *CTH* or Mg-protoporphyin IX monomethylester cyclase; 13, *POR* and *CHLB*/ *CHLL*/ *CHLN* or light-dependent protochlorophyllide oxidoreductase and light-independent protochlorophyllide reductase, respectively; 14, *DVR* or divinyl-protochlorophyllide vinyl reductase; 15, *CHLG* or chlorophyll synthase; 16, *CAO* and *GGR* or chlorophyllide a oxidase and geranylgeranyl

reductase; 17, *CLH* or chlorophyllase, *UPM* or precorrin-2-synthase, *SIRB* or sirohydrochlorin ferrochelatase, *FeC* or ferrochelatase, *HMOX* or heme oxygenase. Full enzyme names and their corresponding GenBank accession numbers are provided in the list of abbreviations.

3.3 Methods

Please refer to section 2.3 for detailed explanation regarding the genome and transcriptome data analyses of the biosynthetic pathways in *Polytomella* spp.

Briefly, characterized *C. reinhardtii* enzymes that are known to function in the tetrapyrrole biosynthetic pathway were used as query sequences to search the *Polytomella* mRNA contigs, and the resulting hits were representative of the enzymes. These hits were checked with non-redundant NCBI database to check for the relevancy of these hits to the tetrapyrrole biosynthetic pathway. No hits to the *Polytomella* transcriptome warranted further investigation into the *Polytomella* genome to check for the presence of these enzymes.

3.4 Results

Genes encoding enzymes involved in steps 1-9 of the tetrapyrrole biosynthetic pathway were present at the transcriptome and genome levels in all four *Polytomella* species. The last enzyme present at both levels is protoporphyrinogen IX oxidase (PPX), which is responsible for the oxidative production of protoporphyrinogen IX from coproporphyrinogen III. After PPX, any attempt to detect genes encoding enzymes at the transcriptome or genome levels was unsuccessful. For instance, step 10 of the pathway involves the addition of magnesium to protoporphyrinogen IX, and is catalyzed by a total of six enzymes: CHLD, CHLH1, CHLH2, CHLI1, and CHLI2, all of which belong to a Mg-chelatase gene family, and GUN4, a Mg-chelatase activating protein (Castelfranco *et al.* 1979; Alberti *et al.* 1995; Willows *et al.* 1996; Larkin *et al.* 2003). All of these genes are missing at the transcriptome and genome levels in the four *Polytomella* species.

Step 11 involves the esterification of the Mg-protoporphyrin IX from the previous step to make Mg-protoporphyrin IX monoethyl ester, a step catalyzed by PPMT, a Mg-

protoporphyrin IX methyltransferase (Figure 6; Beale 2009). The gene is once again absent from any transcriptome or genome contigs. The ester produced in step 11 of the tetrapyrrole biosynthetic pathway is then converted into divinyl protochlorophyllide a via two Mg-protoporphyrin IX monomethylester cyclase enzymes (CRD1 and CRTH1 a/b) (Moseley *et al.* 2000, 2002). These genes were undetected upon analyses of genome and transcriptome data.

Light-dependent protochlorophyllide oxidoreductase, or POR, is responsible for catalyzing step 13, which consists of converting divinyl protochlorophyllide (a) into divinyl chlorophyllide (a) as well as transforming protochlorophyllide (a) into chlorophyllide (a) (see Figure 6, Begley and Young 1989). POR is missing from the transcriptome and genome of all four species.

Light-independent reduction of divinyl protochlorophyllide (a) into protochlorophyllide (a) and divinyl chlorophyllide (a) into chlorophyllide (a) uses enzymes that are encoded within the plastid genome in *C. reinhardtii* (Choquet *et al.* 1992; Huang and Liu 1992; Li *et al.* 1993; Liu *et al.* 1993). The genes encoding these enzymes, namely CHLB, CHLL, and CHLN, enzymes that belong to the light-independent protochlorophyllide reductase class of enzymes, are missing from the transcriptome and genome datasets for the four *Polytomella* species.

Divinyl (proto) chlorophyllide vinyl reductase (or DVR1) catalyzes the last steps involved in chlorophyll production from chlorophyllide (a/b) (Nagata *et al.* 2005). Similarly, it is missing in all species.

Chlorophyllide (a) and (b) are converted into chlorophylls (a) or (b) respectively, via the work of chlorophyll synthase. The gene encoding chlorophyll synthase is also absent for the four *Polytomella* species.

The last and final step, involves the use of chlorophyllide (a) oxidase (CAO) and geranylgeranyl reductase (GGR) for the conversion of chlorophyllide (a) to chlorophyllide (b), to make chlorophyll (b), both of which are missing from the *Polytomella* species (Figure 6; Tanaka *et al.* 1998).

Genes encoding enzymes responsible for heme- and chlorophyll-degradation enzymes, heme oxygenase 1 (HMOX1) and chlorophyllase (CLH) respectively, are absent in all 4 species (Hörtensteiner *et al.* 1998; Vavilin and Vermaas 2007). In contrast, the gene for heme oxygenase 2 (HMOX2) is present in the *Polytomella* species. Similarly, genes for heme and sirohemes synthetic enzymes, precorrin-2 synthase and precorrin-2 dehydrogenase (UMP/SUMT1), sirohydrochlorin ferrochelatase (SIRB), and ferrochelatase (FeC/ HEM15) are all present within the *Polytomella* species (Raux *et al.* 2003; Dailey 1990).

Table 5: Tetrapyrrole biosynthetic pathway enzymes, where ✓ or ✗ represents the presence or absence of enzymes, respectively. Enzymes absent at the transcriptome were checked at the genome level as well, in which case ✗/✗ denotes the absence of the enzyme at both levels.

Step	ENZYMES	*P. capuana*	*P. magna*	*P. parva*	*P. piriformis*
1	GTS1	✓	✓	✓	✓
	GTS2	✓	✓	✓	✓
2	HEMA/GTR	✓	✓	✓	✓
	HEMA	✓	✓	✓	✓
3	GSA	✓	✓	✓	✓
	GSA/GSAT	✓	✓	✓	✓
4	ALAD	✓	✓	✓	✓
5	PBGD	✓	✓	✓	✓
6	UROS	✓	✓	✓	✓
7	UROD1	✓	✓	✓	✓
	UROD2	✓	✓	✓	✓
	UROD3	✓	✓	✓	✓
8	CPX1	✓	✓	✓	✓
	CPX2	✓	✓	✓	✓
9	PPX	✓	✓	✓	✓
10	CHLD	✗/✗	✗/✗	✗/✗	✗/✗
	CHLH1	✗/✗	✗/✗	✗/✗	✗/✗
	CHLH2	✗/✗	✗/✗	✗/✗	✗/✗
	CHLI1	✗/✗	✗/✗	✗/✗	✗/✗
	CHLI2	✗/✗	✗/✗	✗/✗	✗/✗
	GUN4	✗/✗	✗/✗	✗/✗	✗/✗
11	PPMT	✗/✗	✗/✗	✗/✗	✗/✗
12	CRD1	✗/✗	✗/✗	✗/✗	✗/✗
	CTH1	✗/✗	✗/✗	✗/✗	✗/✗
	CTH1b	✗/✗	✗/✗	✗/✗	✗/✗
13	POR	✗/✗	✗/✗	✗/✗	✗/✗

	CHLB	✗/✗	✗/✗	✗/✗	✗/✗
	CHLL	✗/✗	✗/✗	✗/✗	✗/✗
	CHLN	✗/✗	✗/✗	✗/✗	✗/✗
14	DVR1	✗/✗	✗/✗	✗/✗	✗/✗
15	CHLG	✗/✗	✗/✗	✗/✗	✗/✗
16	CAO	✗/✗	✗/✗	✗/✗	✗/✗
	GGR	✗/✗	✗/✗	✗/✗	✗/✗
17	CLH	✗/✗	✗/✗	✗/✗	✗/✗
	UMP/SUMT1	✓	✓	✓	✓
	SIRB	✓	✓	✓	✓
	FeC/HEM15	✓	✓	✓	✓
	HMOX1	✗/✗	✗/✗	✗/✗	✗/✗
	HMOX2	✓	✓	✓	✓

3.5 Discussion

The results demonstrate that tetrapyrrole biosynthesis is partially occurring in the plastids of *Polytomella* species. The initial steps leading onto the second branch point protoporphyrinogen IX (from which heme biosynthesis begins) are present at the genome and transcriptome levels. From then on, the pathway is degraded, erasing genes and transcripts that lead onto chlorophyll a and b biosynthesis.

It is no surprise that in a nonphotosynthetic species, such as *Polytomella*, chlorophyll biosynthesis has been eliminated. All missing proteins are involved in magnesium chelation to form chlorophyll, chlorophyll regulation, or greening of etiolated tissues, functions that are no longer needed for *Polytomella*. Even chlorophyll-degradatory enzymes have been eliminated, mirroring the absence of chlorophyll and any related processes. Chloroplast-encoded light-independent protochlorophyllide reductases were absent within the *Polytomella* transcriptome and genome datasets, supporting the loss of the plastid genome within these species (Smith and Lee 2014).

In contrast, enzymes responsible for heme and siroheme biosynthesis are retained, reflecting the crucial importance of this pathway to these species, and plants and algae as a whole. In fact, some of the enzymes involved in tetrapyrrole biosynthesis were shown to function within *Polytomella* previously, using functional complementation in mutant *Escherichia coli* strains (Atteia *et al.* 2005). Nuclear-encoded, plastid-targeted GSAT, PPO, and FeC are present within the *Polytomella* plastid, similar to other green algal yet

parasitic species, such as *Helicosporidium* and *Prototheca wickerhamii* (de Koning and Keeling 2004; Borza *et al.* 2005; Atteia *et al.* 2005). All three green-algal species encode GSAT and thus, use the glutamate pathway to heme biosynthesis, in contrast to the apicomplexan parasite *Plasmodium falciparum*, which uses the Shemin pathway, depending on its mitochondrial ALA synthase to form ALA (Ralph *et al.* 2004).

3.6 Abbreviations

Tetrapyrrole biosynthetic enzymes (in order of appearance within the pathway)—

1. GTS1/2: Glutamyl-tRNA synthetase
2. GTR/ HEMA: Glutamyl-tRNA reductase
3. GSAT/ GSA: Glutamate 1-semialdehyde aminotransferase (Glutamate 1-semialdehyde aminomutase)
4. PBGS/ ALAD: Porphobilinogen synthase (ALA dehydratase)
5. HMBS/ PBGD: Hydroxymethylbilane synthase (PBG deaminase)
6. UROS: Uroporphyrinogen III synthase
7. UROD1/2/3: Uroporphyrinogen III decarboxylase
8. CPX1/2: Coporphyrinogen III oxidative decarboxylase
9. PPX/ PPO: Protoporphyrinogen IX oxidase
10. CHLD/CHLH1/CHLH2/CHLI1/CHLI2: Mg-chelatase
    GUN4: Mg chelatase-activating protein
11. PPMT: Mg- protoporphyrin IX methyltransferase
12. CRD1/CTH1: Mg-protoporphyrin IX monomethylester cyclase
13. POR: Light-dependent protochlorophyllide oxidoreductase
    CHLB/ CHLL/ CHLN: Light-independent protochlorophyllide reductase
14. DVR: Divinyl (proto) chlorophyllide vinyl reductase
15. CHS (CHLG): Chlorophyll synthase
16. CAO: Chlorophyllide a oxidase
    GGR: geranylgeranyl reductase
17. CLH: Chlorophyllase
    UPM: Precorrin-2 synthase
    SIRB: sirohydrochlorin ferrochelatase

FeC: Ferrochelatase

HMOX1/2: Heme oxygenase

3.7 References

Alberti, M., Burke, D.H., and Hearst, J.E. 1995. Structure and sequence of the photosynthesis gene cluster. In: Anoxygenic Photosynthetic Bacteria (R.E. Blankenship, M.T. Madigan, and C.E. Bauer, Eds.), pp. 1083-1106. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Atteia, A., van Lis, R., and Beale, S.I. 2005. Enzymes of the heme biosynthetic pathway in the nonphotosynthetic alga *Polytomella* sp. Eukaryot Cell 4: 2087-2097.

Beale, S.I., and Castelfranco, P.A. 1974. The biosynthesis of δ-aminolevulinic acid in higher plants. II. Formation of ^{14}C- δ-aminolevulinic acid from labeled precursors in greening plant tissues. Plant Physiol 53: 297-303.

Beale, S.I., Gough, S.P., and Granick, S. 1975. The biosynthesis of δ-aminolevulinic acid from the intact carbon skeleton of glutamic acid in greening barley. Proc Natl Acad Sci USA 72: 2719-2723.

Begley, T.P., and Young, H. 1989. Protochlorophyllide reductase. 1. Determination of the regiochemistry and the stereochemistry of the reduction of protochlorophyllide to chlorophyllide. J Am Chem Soc 111: 3095-3096.

Borza, T., Popescu, C.E., and Lee, R.W. 2005. Multiple metabolic roles for the nonphotosynthetic plastid of the green alga *Prototheca wickerhamii*. Eukaryo Cell 4: 253-261.

Castelfranco, P.A., Weinstein, J.D., Schwarcz, S., Pardo, A.D., and Wezelman, B.E. 1979. The Mg insertion step in chlorophyll biosynthesis. Arch Biochem Biophys 192: 592-598.

Choquet, Y., Rahire, M., Girard-Bascou, J., Erikson, J., and Rochaix, J.D. 1992. A chloroplast gene is required for the light-independent accumulation of chlorophyll in *Chlamydomonas reinhardtii*. EMBO J 11:1697-1704.

Dailey, H.A. 1990. Conversion of coproporphyrinogen to protoheme in higher eukaryotes and bacteria: terminal three enzymes. In: Biosynthesis of Heme and Chlorophylls (H.A.Dailey, Ed.) pp. 123-161. McGraw-Hill, New York.

de Koning, A.P., and Keeling, P.J. 2004. Nucleus-encoded genes for plastid targeted proteins in Helicosporidium: functional diversity of a cryptic plastid in a parasitic alga. Eukaryot Cell 3: 1198-1205.

Gibson, K.D., Laver, W.G., and Neuberger, A. 1958. Formation of δ-aminolevulinic acid in vitro from succinyl-coenzyme a and glycine. Biochem J 70: 71-81.

Hörtensteiner, S., Wüthrich, K.L., Matile, P., Ongania, K.H., and Kräutler, B. 1998. The key step in chlorophyll breakdown in higher plants; cleavage of pheophorbide a macrocycle by a monooxygenase. J Biol Chem 273: 15335-15339.

Huang, C., and Liu, X.Q. 1992. Nucleotide sequence of the frxC, petB and trnL genes in the chloroplast genome of *Chlamydomonas reinhardtii*. Plant Mol Biol 18: 985-988.

Kannangara, C.G., and Gough, S.P. 1978. Biosynthesis of δ-aminolevulinate in greening barley leaves: glutamate 1-semialdehyde aminotransferase. Carlsberg Res Commun 43: 185-194.

Kannangara, C.G., Gough, S.P., Bruyant, P., Hoober, J.K., Kahn, A., and von Wettstein, D. 1988. tRNA(Glu) as a cofactor in delta-aminolevulinate biosynthesis: steps that regulate chlorophyll synthesis. Trends Biochem Sci 13: 139-143.

Kannangara, C.G., Gough, S.P., Oliver, R.P., and Rasmussen, S.K. 1984. Biosynthesis of δ-aminolevulinate in greening barley leaves. VI. Activation of glutamate by ligation to RNA. Carlsberg Res Commun 49: 417-437.

Kikuchi, G., Kumar, A., Talmage, P., and Shemin, D. 1958. The enzymatic synthesis of δ-aminolevulinic acid. J Biol Chem 233: 1214-1219.

Larkin, R.M., Alonso, J.M., Ecker, J.R., and Chory, J. 2003. GUN4, a regulator of chlorophyll synthesis and intracellular signaling. Science 209: 902-906.

Li, J., Goldschmidt-Clermont, M., and Timko, M.P. 1993. Chloroplast-encoded chlB is required for light-independent protochlorophyllide reductase activity in *Chlamydomonas reinhardtii*. Plant Cell 5: 1817-1829.

Liu, X.Q., Xu, H., and Huang, C.1993. Chloroplast chlB is required for light-independent chlorophyll accumulation in *Chlamydomonas reinhardtii*. Plant Mol Biol 23: 297-308.

Meller, E., Belkin, S., and Harel, E. 1975. The biosynthesis of δ-aminolevulinic acid in greening maize leaves. Phytochemistry 14: 2399-2402.

Moseley, J.L., Quinn, J., Eriksson, M., and Merchant, S. 2000. The Crd1 gene encodes a putative di-iron enzyme for photosystem I accumulation in copper deficiency and hypoxia in *Chlamydomonas reinhardtii*. EMBO J 19: 2139-2151.

Moseley, J.L., Page, M.D., Adler, N.P., Eriksson, M., Quinn, J., Soto, F., Theg, S.M., Hippler, M., and Merchant, S. 2002. Reciprocal expression of two candidate di-iron enzymes affecting photosystem I and light-harvesting complex accumulation. Plant Cell 14: 673-688.

Nagata, N., Tanaka, R., Satoh, S., and Tanaka, A. 2005. Identification of a vinyl reductase gene for chlorophyll synthesis in *Arabidopsis thaliana* and implications for the evolution of *Prochlorococcus* species. Plant Cell 17: 233-240.

Panek, H., and O'Brian, M.R. 2002. A whole genome view of prokaryotic haem biosynthesis. Microbiology 148: 2273-2282.

Papenbrock, J., and Grimm, B. 2001. Regulatory network of tetrapyrrole biosynthesis—studies of intracellular signaling involved in metabolic and developmental control of plastids. Planta 213: 667-681.

Pontoppidan, B., and Kannangara, C.G. 1994. Purification and partial characterization of barley glutamyl-tRNA^{Glu} reductase, the enzyme that directs glutamate to chlorophyll biosynthesis. Eur J Biochem 225:529-537.

Ralph, S.A., van Dooren, G.G., Waller, R.F., Crawford, M.J., Fraunholz, M.J., Foth, B.J., Tonkin, D.S.R., and McFadden, G.I. 2004. Tropical infectious diseases: metabolic maps and functions of the *Plasmodium falciparum* apicoplast. Nat Rev Microbiol 2: 203-216.

Raux, E., Leech, H.K., Beck, R., Schubert, H.L., Santander, P.J., Roessner, C.A, Scott, A.I., Martens, J.H., Jahn, D., Thermes, C., Rambach, A., and Warren, M.J. 2003. Identification and functional analysis of enzymes required for precorrin-2-synthase dehydrogenation and metal ion insertion in the biosynthesis of sirohaem and cobalamin in *Bacillus megaterium*. Biochem J 370: 505-516.

Sato, S., Clough, B., Coates, L., and Wilson, R.J.M.I. 2004. Enzymes for heme biosynthesis are found in both the mitochondrion and plastid of the malaria parasite *Plasmodium falciparum*. Protist 155: 117-125.

Smith, D.R., and Lee, R.W. 2014. A plastid without a genome: evidence from the nonphotosynthetic green algal genus *Polytomella*. Plant Physiol 164: 1812-1819.

Tanaka, A., Ito, H., Tanaka, R., Tanaka, N.K., Yoshida, K., and Okada, K. 1998. Chlorophyll a oxygenase (CAO) is involved in chlorophyll b formation from chlorophyll a. Proc Natl Acad Sci USA 95: 12719-12723.

Varadharajan, S., Dhanasekaran, S., Bonday, Z.Q., Rangarajan, P.N., and Padmanaban, G. 2002. Involvement of 5-aminolevulinate synthase encoded by the parasite gene in de novo haem synthesis by *Plasmodium falciparum*. Biochem J 367: 321-327.

Vavilin, D., and Vermaas, W. 2007. Continuous chlorophyll degradation accompanied by chlorophyllide and phytol reutilization for chlorophyll synthesis in *Synchocystis* sp. PCC 6803. Biochim Biophys Acta 1767: 920-929.

Willows, R.D., Gibson, L.C.D., Kannangara, C.G., Hunter, C.N., and von Wettstein, D. 1996. Three separate proteins constitute the magnesium chelatase of *Rhodobacter sphaeroides*. Eur J Biochem 235: 438-443.

Chapter 4

4 Amino Acids Biosynthetic Pathway

4.1 Introduction: Amino acid pathways at a glance.

A multitude of amino acids and their respective metabolic pathways are present not only to serve as building blocks for proteins, but also to provide various secondary metabolites necessary for an organism's survival. Some amino acid biosynthetic pathways are specifically induced under stress conditions in various organisms, whereas others are expressed in phototrophs only, making them effective herbicide targets (Mourad and King 1992; Ott *et al.* 1996; Fabro *et al.* 2004; Verslues and Bray 2006). Still others are metabolized only within parasites, providing effective chemotherapeutic targets for treatment of specific diseases (Coggins *et al.* 2003; Ralph *et al.* 2004). Some amino acids are known as non-essential amino acids; these amino acids are derived from and synthesized within the organism *de novo*, whereas essential amino acids are those which have to be supplemented through diet and supplements and cannot be synthesized in certain organisms, mostly animals including humans (Ferreira *et al.* 2005). For instance, branched amino acids are essential for insects and have to be obtained from external sources (Akman Gunduz and Douglas 2009).

Amino acids can be classified into general categories, such as aromatic amino acids (e.g., phenylalanine, tyrosine, and tryptophan), branched chain amino acids (e.g., valine, leucine, and isoleucine), aspartate-derived amino acids (e.g., asparagine, threonine, methionine, and lysine), glutamate-derived amino acids such as arginine, proline, aspartate, alanine, serine, glycine, and others including cysteine and histidine.

In *Chlamydomonas reinhardtii*, and most other photosynthetic green algae, there are 21 genetically encoded amino acids, including the standard 20 amino acids as well as selenocysteine. Some amino acid pathways are particularly noteworthy in organisms as a whole; for instance, in abiotically stressed green plants, proline and arginine biosynthetic enzymes are enhanced, whereas the reverse is true for all other amino acids (Less and Galili 2008). The assimilation of ammonium constitutes an important factor in the

survival and growth of amino acid- synthesizing organisms (Kropat *et al.* 2011). Ammonium is an inorganic nitrogen source within cells and is used along with carbon skeletons for the assimilation of over 21 amino acids. Whereas ammonia (NH_3) transport across the membrane is passive under basic conditions (high pH levels), ammonium (NH_4^+) transport is regulated and requires specific anionic channels for its translocation across membranes. The uptake of nitrate is composed of two reduction steps and their respective transport steps. Nitrate is transported into the cytosol where it is reduced by nitrate reductase into nitrite, which is transported into the plastid and reduced to ammonium via nitrite reductase. The ammonium is then ligated onto carbon skeletons via the glutamine synthase/ glutamine oxoglutarate amidotransferase (GS/GOGAT) cycle for the formation of various amino acids and proteins (Hoff *et al.* 1994; Crawford 1995).

Aromatic amino acid synthesis within photosynthetic land plants and green algae proceeds via the shikimate pathway, solely named so for the intermediate shikimate, which is indispensible for the organism's survival and growth (Dewick 1998; Knaggs 2001; Knaggs 2003). The three aromatic compounds, namely phenylalanine, tyrosine, and tryptophan, are synthesized within the plastid using nucleus-encoded, plastid-targeted enzymes, involved in converting chorismate (the common precursor) into the 3 different aromatic amino acids. They constitute less than 10% of total amino acid concentration within photosynthetic organisms and can be isolated using spectrophotometry—this is because aromatic amino acids absorb UV wavelengths of 260 to 280 nm (Layne 1957). Aromatic amino acids can be synthesized *de novo* in plants, algae, fungi, and various other microorganisms, but are "essential" in animals and have to be supplemented through dietary sources (Herrmann and Weaver 1999; Leuchtenberger *et al.* 2005).

Similarly, lysine, methionine, threonine, and asparagine are synthesized from aspartate via the aspartate-derived amino acid biosynthetic pathway, which is present in most plants and algae, fungi, and bacteria, but is absent in animals, meaning that the four amino acids are essential to animals. Animal feed is usually deficient in one or more of these enzymes, making it essential to supplement the feed with millions of dollars worth of amino acid supplements. As a result, it would be greatly beneficial and economically

beneficial to genetically enhance plants' own biomass with increased levels of these amino acids for animal and livestock feed purposes (Mueller and Huebner 2003).

The biosynthesis of amino acids involves the import of nitrite from the cytosol and its conversion into ammonia within the plastid, after which it is incorporated into the GS/GOGAT cycle for the formation of glutamine and glutamate (Figure 7). Glutamate is converted into oxaloacetate to be further converted into α-ketoglutarate to form proline and arginine, or into aspartate for the formation of threonine, methionine, lysine, and asparagine. Threonine can further help, via its conversion into 2-oxobutanoate and acetyl-CoA and the addition of pyruvate, in the formation of branched amino acids: valine, leucine, and isoleucine. Glutamate can also form 3-phosphoglycerate, which is the common precursor for serine and glycine. Methionine derived from aspartate in the previous steps can be converted into homocysteine for the formation of cysteine and alanine with desulfuration reactions. The oxidative pentose phosphate pathway is responsible for supplementing phosphoenolpyruvate (PEP) and erythrose-4-phosphate (E4P), which are the precursors for the shikimate pathway that ends with the common precursor chorismate responsible for the formation of aromatic amino acids.

The biosynthesis of branched chain amino acids, valine, leucine, and isoleucine, is not only studied in terms of its metabolic roles within synthesizing organisms but also for its regulatory circuitry and importance for non-synthesizing animals, including humans (Pátek 2007). It has been the target of developing competitive inhibitory substrates for anti-tuberculosis medication and herbicides targeting plant pathways that are absent in the animals that consume these plants (Grandoni *et al.* 1998; Wang *et al.* 2005). Isoleucine and valine biosynthesis share four enzymes in total and are synthesized in parallel, whereas leucine biosynthesis branches from one of valine's intermediates, 3-methyl-2-oxobutanoate in a four-step reaction ending with leucine.

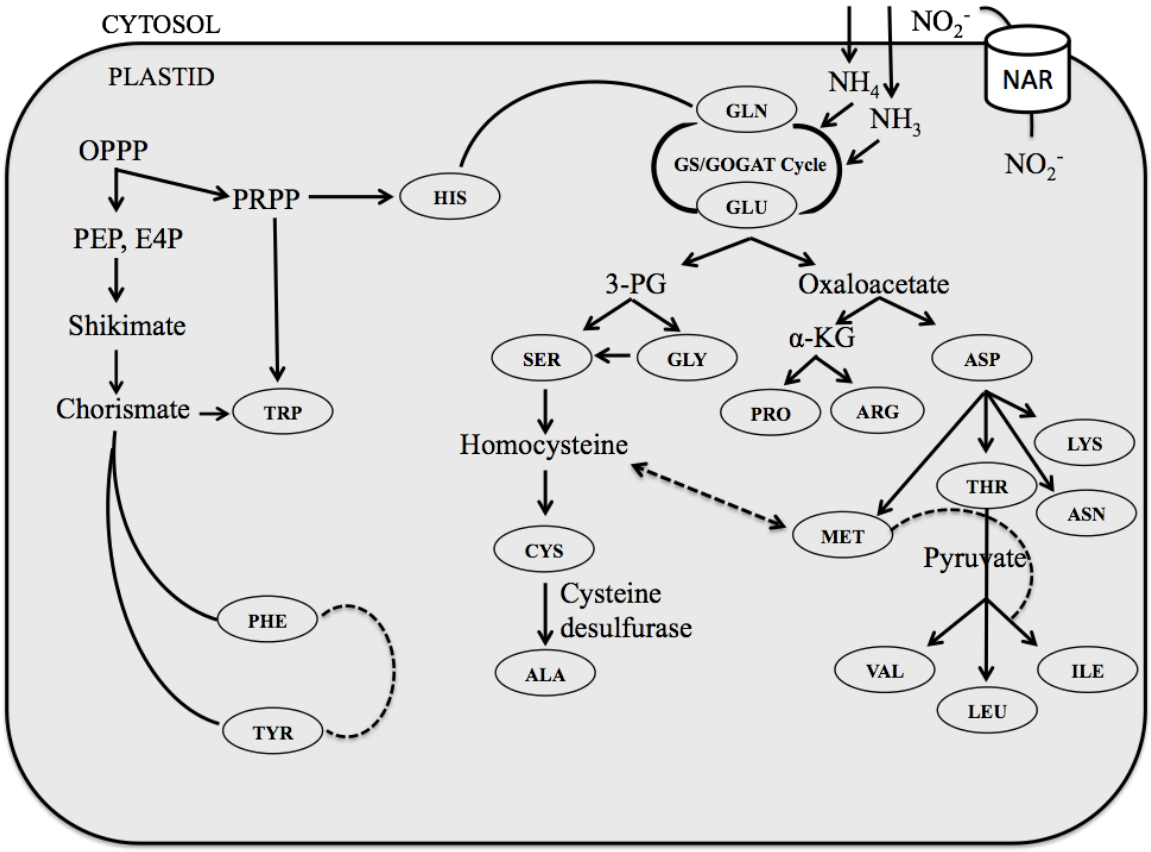


Figure 7: Summary of amino acid biosynthetic pathways predicted to occur within the *Polytomella* plastid. Figures 8-11 show the biosynthetic pathways in detail along with their associated enzymes.

Arginine and proline biosynthesis occurs in a similarly coupled manner to branched chain amino acids, whereby proline can be synthesized either via glutamate conversion into cyclized non-acetylated glutamate-5-semialdehyde (GSA) or via ornithine (one of arginine pathway intermediates) conversion into GSA. The study of arginine and proline biosynthesis is important for understanding key producers of plant alkaloids and animal-feeding deterrants, as well as characterizing proline's roles within green algae in terms of combatting biotic and abiotic stress, as shown in land plants (Rosenthal 1982; Facchini 2001; Szekely *et al.* 2008; Sharma and Verslues 2010; Fabro *et al.* 2004).

4.2 Amino acid pathways in detail.

Ammonia assimilation into glutamine and glutamate involves the use of three enzymes: glutamine synthetase (GS), glutamate synthetase (with its many isoforms, GOGAT, GSF1, GSN1), and catabolic glutamate dehydrogenase (GDH) (Figure 7; Miflin and Lea 1976; Fischer and Klein 1988). In the first step, ammonia and glutamate form the amide group of glutamine in an ATP-dependent reaction catalyzed by glutamine synthetase 1 (GS1—located in the cystosol) and glutamine synthetase 2 (GS2—located in the plastid). There are a total of four genes encoding glutamine synthetases. *GLN1* and *GLN4* encode GS1, whereas *GLN2* and *GLN3* encode the plastidic isoform GS2. The amide group of glutamine is then transferred to α-ketoglutamate to form 2 glutamate molecules. This reaction is catalyzed by glutamate synthetase (GOGAT), NADH-GOGAT (which is encoded by *GSN1*), and Fd-GOGAT (encoded by *GSF1*). Glutamate dehydrogenase (GDH) serves a catabolic function whereby it is involved in protein turnover and degradation (Cullimore and Sims 1981). Several studies suggest that GDH is localized to the plastid (Fischer and Klein 1988).

For aromatic amino acid biosynthesis pathway, phosphoenolpyruvate (PEP) from the glycolytic pathway and erythrose-4-phosphate of the pentose phosphate pathway come together in a condensation reaction to form the first shikimate intermediate, 3-deoxy-d-arabino-heptulosonate-7-phosphate (DAHP). This is catalyzed by the enzyme 3-deoxy-d-arabino-heptulsonate-7-phosphate synthase (DAHPS or SHKA1) (Figure 8; Srinivasan and Sprinson 1959; Herrmann and Weaver 1999). 3-deoxy-d-arabino-heptulosonate-7-phosphate is converted to 3-dehydroquianate by catalytic 3-dehydroquianate synthase (DHQS). A two-step reaction involving the conversion of 3-dehydroquianate to 3-dehydro shikimate, and eventually to shikimate is catalyzed by the bifunctional enzyme 3-dehydroquianate dehydrogenase/ shikimate-5-dehydrogenase (DHQ/SDH or SHKD1). Figure 8 depicts the next step in which shikimate is phosphorylated into shikimate-3-phosphate via shikimate kinase (or SHKF1). Shikimate-3-phosphate is further converted into 5-enolpyruvylshikimate-3-phosphate and leads to chorismate, via the catalytic work of 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) and chorismate synthase (CS), respectively. Chorismate is the main precursor of all three amino acids, after which

tryptophan biosynthesis branches out from the tyrosine and phenylalanine biosynthetic pathways (Figure 8). DAHPS forms a polypeptide complex with chorismate mutase, a reaction responsible for the conversion of chorismate to prephenate in the phenylalanine and tyrosine branch of this pathway (Ikeda 2006). Prephenate is further converted into arogenate, via prephenate aminotransferase (PAT), to make phenylalanine and tyrosine, by arogenate dehydratase (ADT) and arogenate/prephenate dehydrogenase (AGD1), respectively (Figure 8). The tryptophan biosynthetic pathway involves the transamination and aromatization of chorismate into anthranilate, via anthranilate synthase (ANS/ASB) (Pittard 1996; Nichols 1996). Anthranilate, with the addition of 5-phosphoribosyl pyrophosphate is converted into phosphoribosylanthranilate by phosphoribosyl anthranilate transferase (PRT). Phosphoribosylanthranilate is converted into indole in a three-step reaction catalyzed by phosphoribosylanthranilate isomerase (PAI), indole-3-glycerol phosphate synthase (IGS), and tryptophan synthase α-subunit (TSA), from which indole is converted into tryptophan by tryptophan synthase β-subunit (TSB or MAA7) (Figure 8).

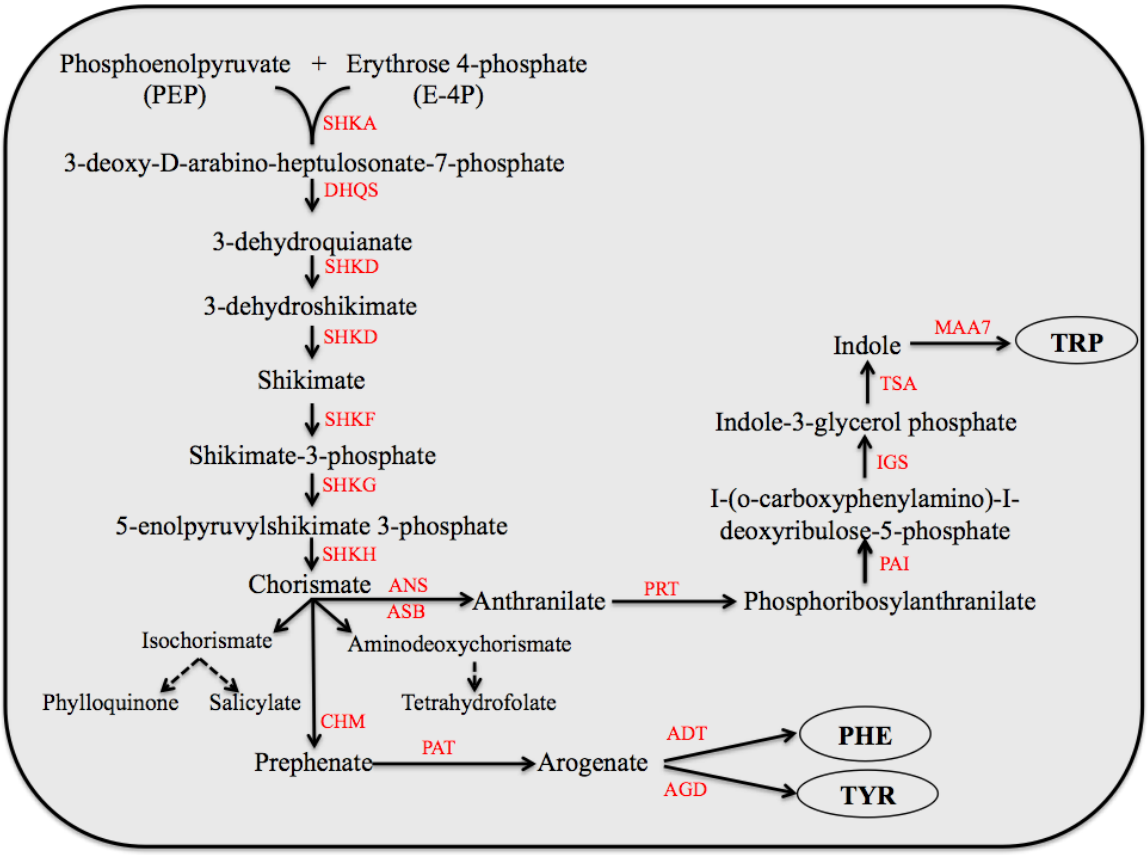


Figure 8: Aromatic amino acid biosynthetic pathway. Chorismate represents the main intermediate from which aromatic amino acids branch. Phenylalanine and tyrosine are synthesized from chorismate to prephenate conversion, whereas tryptophan is synthesized from chorismate to indole conversion. *SHKA*, 3-deoxy-D-arabino-heptylosonate-7-phosphate synthase; *DHQS*, 3-dehydroquinate synthase; *SHKD1*, bifunctional dehydroquinate dehydratase-shikimate: NADP oxidoreductase; *SHKF1*, shikimate kinase; *SHKG1*, 5-enolpyruvylshikimate-3-phosphate synthase; *SHKH1*, chorismate synthase; *CHM1*, chorismate mutase; *AST*, aspartate aminotransferase; *PAT*, phosphate acetyltransferase; *PRD1*, prephenate dehyrdratase; *AGD1*, prephenate dehydrogenase; *AAH1*, aromatic amino acid hydroxylase; *ANS1*, anthranilate synthase α-subunit; *ASB1*, anthranilate synthase, β-subunit; *PRT1*, anthranilate phosphoribosyltransferase; *ASB2*, phosphoribosylanthranilate isomerase; *IGS1*, indole-3-glycerol-phosphate synthase; *TSA*, tryptophan synthetase, α-subunit; *MAA7*, tryptophan synthase, β-subunit. Full

enzyme names are listed in Section 4.6, and accession numbers are provided in the list of abbreviations.

Lysine, methionine, threonine, and asparagine are four of the eight essential amino acids that humans and other animals are unable to synthesize, and are all derived from the common precursor L-aspartate—the reason they are called the aspartate-derived amino acid group. The pathway begins with the transamination of oxaloacetate and glutamate by aspartate aminotransferases (AST1-5) producing aspartate and α-ketoglutarate. Direct amidation of aspartate results in the formation of asparagine, via the catalytic work of asparagine synthetase (ASNS), which has been shown to occur in the cytosol of green algae and land plants. In *C. reinhardtii*, ASNS lacks an N-terminal plastid-targeted sequence, which suggests the enzyme acts in the cytosol as opposed to the plastid. Aspartate can also be phosphorylated to form aspartyl-4-phosphate via aspartate kinase, one of the most regulated enzymes of this pathway. The latter is then dehydrogenated by aspartyl semialdehyde dehydrogenase (ASSD) to produce aspartate-4-semiladehyde, the last common precursor of the other three amino acids: lysine, methionine, and threonine (Figure 9). Aspartate-4-semiladehyde is then reduced to homoserine, which is phosphorylated to O-phospho-L-homoserine, using homoserine dehydrogenase (HSD) and homoserine kinase (HSK), respectively. O-phospho-L-homoserine is dephosphorylated and its hydroxyl groups rearranged to result in threonine via threonine kinase (THS; Figure 9). Threonine degradation can be facilitated by threonine deaminase and/or threonine aldolase (THA). Threonine deaminase (THD) is responsible for the deamination of threonine to 2-oxobutanoate, for the start of branched amino acid biosynthetic pathway, and specifically for the formation of isoleucine.

O-phospho-L-homoserine, with the addition of cysteine, can also result in cystathionine by cystathionine –γ-synthase (CGS) in a step committed to methionine formation. The reduction of cystathionine to L-homocysteine by methionine synthase requires methionine synthase, of which there are two types: cobalamin (vitamin B_{12})-dependent methionine synthase (METH 1 and 2 isoforms) and cobalamain (vitamin B_{12})-independent methionine synthase (METE). Methionine can be degraded into 2-

oxobutanoate and/or S-adenosyl-L-methionine by methionine-γ-lyase (MGL) and S-adenosylmethionine synthetase (SAM), respectively (Figure 9).

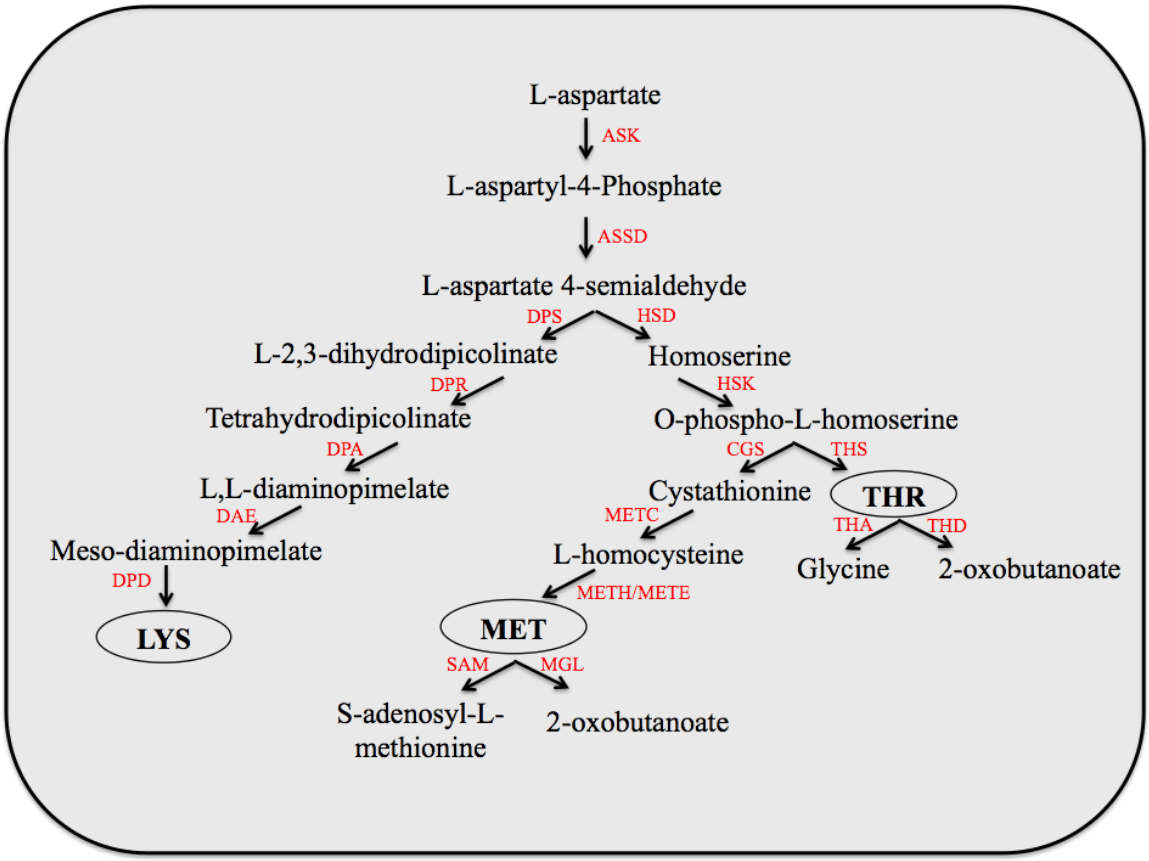


Figure 9: Aspartate-derived amino acid biosynthetic pathway. L-aspartate-4-semiladehyde represents the branch point to lysine, or methionine and threonine. *ASK1*, aspartate kinase; *ASSD*, aspartate semialdehyde dehydrogenase; *HSD*, homoserine dehydrogenase; *HSK*, homoserine kinase; *THS*, threonine synthase; *THD*, threonine dehydratase; *THA*, threonine aldolase; *CGS*, cystathionine-γ-synthase; *METC*, cystathionine-β-lyase; *METE*, cobalamin-independent methionine synthase; *METH*, cobalamin-dependent methionine synthase; *SAM*, S-adenosylmethionine synthetase; *MGL*, methionine-γ-lyase; *DPS*, dihydropicolinate synthase; *DPR*, dihydropicolinate reductase; *DPA1*, L, L-diaminopimelate aminotransferase; *DAE1*, diaminopimelate epimerase; *DPD1*, diaminopimelate decarboxylase. Full enzyme names are listed in Section 4.6, and GenBank accession numbers are provided in the list of abbreviations.

The condensation of L-aspartate-4-semiladehyde into L-2,3-dihydrodipicolinate by dihydrodipicolinate synthase (DPS) represents the first step for lysine formation. The following steps (shown in figure 9 above) involve the reduction of L-2,3-dihydrodipicolinate and the transamination of tetrahydrodipicolinate and epimerization of L,L-diaminopimelate to form meso-diaminopimelate leading to the decarboxylation of the latter to form lysine by diaminopimelate decarboxylase (DPD).

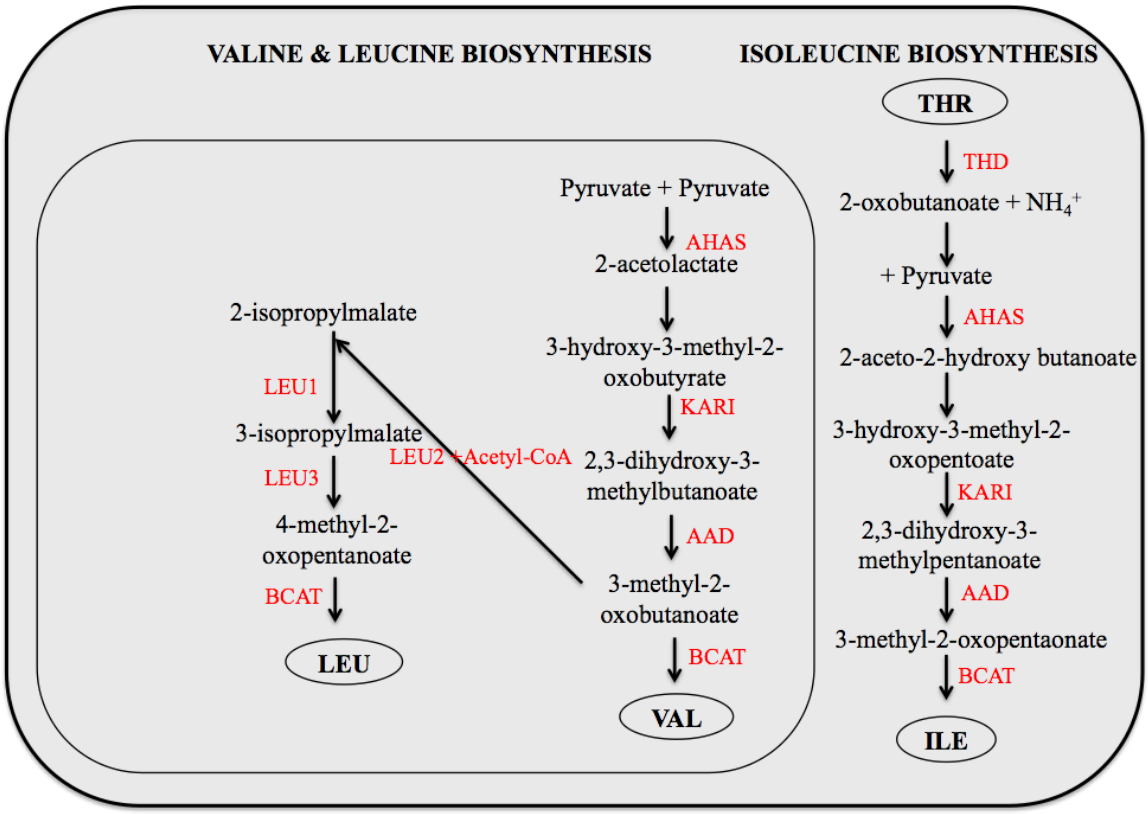


Figure 10: Branched chain amino acid biosynthetic pathway. Parallel pathways with four shared enzymes result in either isoleucine or valine. Leucine biosynthesis continues on from valine biosynthesis pathway. *THD1*, threonine deaminase; *AHAS*, acetolactate synthase, small subunit; *KARI*, acetohydroxyacid isomeroreductase; *AAD*, dihydroxyacid dehydratase; *BCAT*, branched-chain aminotransferase; *LEU1L*, 3-isopropylmalate dehydratase, large subunit; *LEU1S*, isopropylmalate dehydratase, small subunit; *LEU2*, 2-isopropylmalate synthase; *LEU3*, 3-isopropylmalate dehydrogenase. Full enzyme names are provided in Section 4.6, along with their GenBank accession numbers in the list of abbreviations.

Branched chain amino acid biosynthetic pathway involves the use of four enzymes in parallel pathways, starting from different precursors and resulting in the different amino acids. Isoleucine biosynthesis begins with the deamination of threonine into 2-oxobutanoate and ammonium, which along with pyruvate results in 2-aceto-2-hydroxy butanoate. The reaction is catalyzed by acetolactate synthase (AHAS) and represents the first enzyme between the valine and isoleucine biosynthetic pathways. 2-aceto-2-hydroxy butanoate is converted into 3-hydroxy-3-methyl-2-oxopentanoate, which is converted into 2,3-dihydroxy-3-methylpentanoate by the action of acetohydroxyacid isomeroreductase (AAI or KARI). 2,3-dihydroxy-3-methylpentanoate is converted into 3-methyl-2-oxopentanoate by dihydroxyacid dehydratase (AAD). In the last step, branched chain aminotransferases (BCAT) convert 3-methyl-2-oxopentanoate into isoleucine.

Valine biosynthesis uses the same set of four enzymes, but starts with two pyruvate molecules condensing to form 2-acetolactate by AHAS. 2-acetolactate is then converted into 2,3-dihydroxy-3-methylbutanoate, by KARI, which is the substrate for DHAD resulting in 3-methyl-2-oxobutanoate. Aminotransferases are then responsible for the conversion of 3-methyl-2-oxobutanoate into valine (Figure 10).

Lastly, as depicted in Figure 10, leucine biosynthesis starts with the addition of acetyl-CoA to 3-methyl-2-oxobutanoate forming 2-isopropylmalate by 2-isopropylmalate synthase (LEU2). 2-isopropylmalate is converted into 3-isopropylmalate, which is transformed into 3-methyl-2-oxopentanoate in two reactions catalyzed by isopropylmalate dehydratase (LEU1) and isopropylmalate dehydrogenase (LEU3). The last step is catalyzed by branched chain aminotransferases (BCAT) to form leucine. In summary, branched chain aminotransferases are responsible for converting 2-oxoacids (such as 3-methyl-2-oxopentanoate (3MOP), 3-methyl-2-oxobutanoate (3MOB), and 4-methyl-2-oxopentanoate (4MOP)) into the specific amino acids as well as the degradation of said enzymes (Singh 1999; Diebold *et al.* 2002; Schuster and Binder 2005; Binder 2010).

The first four steps of arginine and proline biosynthesis (Figure 11) are conserved within plants, and involve either the acetylation of glutamate in the arginine branch or the direct cyclization of glutamate in the proline branch of the pathway (Shargool *et al.* 1988; Thompson 1980; Slocum 2005). The acetylation of glutamate results in N-acetylglutamate-5-semialdehyde, which results in the first committed step to arginine and the formation of ornithine via N-acetylornithine deacetylase (NAOD or AOD) and carbamoylphosphate synthase (CMPS or CMPL). NAOD has been shown to be absent in land plants, with the exception of *A. thaliana*, implying the absence of a linear ornithine biosynthetic pathway in some plants (Shargool *et al.* 1988; Marchler-Bauer *et al.* 2003). In plants that possess it, the NAOD enzyme does not have the plastid-targeting sequence, implying that it does not function in the plastid but rather works in the cytosol (Slocum 2005; Yokota *et al.* 2002). Carbomylation of ornithine results in citrulline, forming arginoinosuccinate which is converted into arginine via argininosuccinate lyase (ARG7).

Proline can be synthesized via glutamate in a direct reaction involving a bifunctional enzyme Δ^1- pyrroline-5-carboxylate synthetase (P5CS), which also possesses γ-glutamyl kinase (PROB 1/2 or γ-GK) activity. The non-acetylation version can be readily cyclized resulting in pyrroline-5-carboxylate, which is reduced to proline. An alternative pathway to proline biosynthesis involves the use of ornithine as a precursor for glutamate-5-semialdehyde generation (Figure 11; Alonso *et al.* 2003).

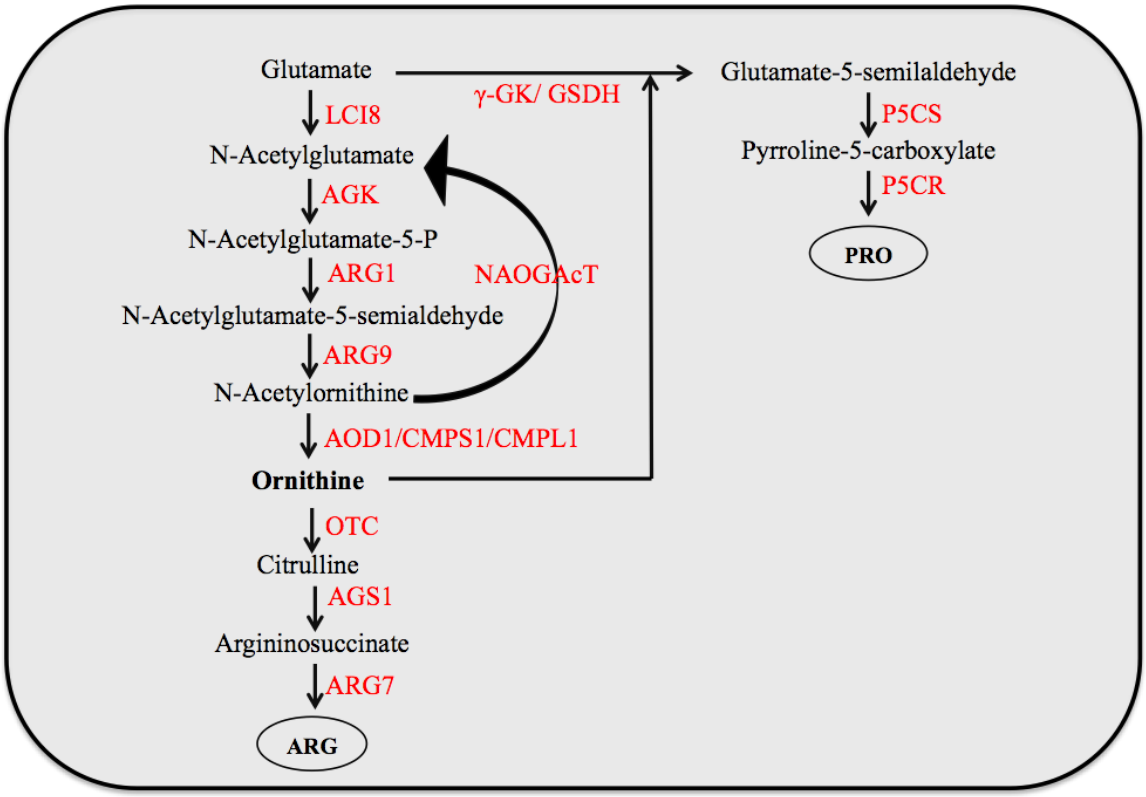


Figure 11: Arginine and proline biosynthetic pathway. Proline could either be synthesized from glutamate or ornithine. *LCI8*, N-acetylglutamate synthase; *AGK*, N-acetyl-L-glutamate kinase; *ARG1*, N-acetyl-γ-glutamyl phosphate reductase; *ARG9*, N-acetylornithine aminotransferase; *AOD1*, N-acetylornithine deacetylase; *CMPS1*, carbamoyl phosphate synthase, small subunit; *CMPL1*, carbamoyl phosphate synthase, large subunit; *OTC*, ornithine carbomyltransferase; *AGS1*, argininosuccinate synthase; *ARG7*, argininosuccinate lyase; *NAOGAct*, N-acetylornithine glutamate aminotransferase; *γ-GK*, γ-glutamyl kinase; *P5CS*, Δ^1-pyrroline-5-carbaoxylate synthetase; *P5CR*, Δ^1-pyrroline-5-carbaoxylate reductase. Full enzyme names are listed in Section 4.6, along with their GenBank accession numbers in the list of abbreviations.

4.3 Methods

Please refer to section 2.3 for detailed explanation of the genomic and transcriptomic data analyses of biosynthetic pathways in *Polytomella* spp.

Enzymes involved in amino acid biosynthesis in the chloroplast of *C. reinhardtii* were used as query sequences to search the *Polytomella* local databases, a total of four databases for each of the *Polytomella* species. This was done at both the transcriptome and genome levels. For enzymes that lacked a transcriptome hit, we used the BLAST function to search the *Polytomella* genome, using the same *C. reinhardtii* mRNA sequences. All hits were then rechecked against the non-redundant NCBI database to confirm that the enzyme was in fact the one we were looking for.

In some cases in which the transcriptome showed a hit whereas the genome data analyses had no hit to the relative enzyme, we followed a different approach by which the *Polytomella* mRNA contig (that was a hit) was taken and used as a query sequence to its corresponding species' DNA data. In some cases, we obtained a better and relevant hit in the DNA, while in others it still failed to give a hit to the enzyme of interest. Again, all hits were then rechecked against the non-redundant NCBI database.

4.4 Results

As shown in Table 6, the three enzymes responsible for ammonia assimilation and the glutamate/glutamine cycle were identified in the transcriptomes and genomes of *P. capuana*, *P. magna*, and *P. piriformis*. Transcripts for glutamate dehydrogenase isoforms GDH1 and GDH2 were not identified in the *P. parva* transcriptome but the genes for these two enzymes were found in its genome. Together, these data indicate that the plastids of *Polytomella* algae have all the enzymes required for ammonia assimilation.

Table 6: Ammonia assimilation enzymes, where ✓ or ✗ represents the presence or absence of the enzyme. ✗/✓ denotes the absence of the enzyme at the transcriptome level, but its presence within the genome data.

NH_3 ASSIMILATION AND GLUTAMATE/ GLUTAMINE BIOSYNTHESIS				
ENZYMES	*P. capuana*	*P. magna*	*P. parva*	*P. piriformis*
GDH1	✓	✓	✗/✓	✓
GDH2	✓	✓	✗/✓	✓
GLN1	✓	✓	✓	✓
GLN2	✓	✓	✓	✓
GLN3	✓	✓	✓	✓
GLN4	✓	✓	✓	✓
GSF1	✓	✓	✓	✓
GSN1	✓	✓	✓	✓

The enzymes leading to chorismate, in the aromatic amino acid biosynthetic pathway, are present in all four *Polytomella* species, at both the transcript and gene levels (Table 7). Aspartate aminotransferases, helping with the conversion of prephenate into arogenate, in the phenylalanine and tyrosine branch of the pathway, possesses all isoforms but AST5 in *P. magna*. Prephenate aminotransferase, also responsible for the conversion of prephenate into arogenate for phenylalanine biosynthesis, is present in all *Polytomella* species. Arogenate dehydratase (ADT or PRD1) and arogenate/prephenate dehydrogenase (AGD) are present at the transcriptome and genome levels (Table 7). Aromatic amino acid hydroxylase (AAH1), responsible for an alternative method of tyrosine biosynthesis from phenylalanine, is absent from the transcriptomes of *P. magna* and *P. parva*, but was identified in their genomes. The AAH1 transcript was found in the other two *Polytomella* species. In the pathway leading to tryptophan, indole-3-glycerol phosphate synthase (MAA4 or IGS) sequence went undetected in the genome and transcriptome data of all *Polytomella* species (Table 7). IGS is responsible for indole-3-glycerol phosphate formation, 2 steps away from tryptophan biosynthesis. In short, the enzymes of leading to phenylalanine, tyrosine, and tryptophan biosynthesis are all present within the *Polytomella* plastids, with the exception of IGS in the tryptophan branch of the pathway.

Table 7: Aromatic amino acid enzymes, where ✓ or ✗ represents the presence or absence of enzymes, respectively. ✗/✓ denotes the absence of the enzyme at the transcriptome level, but its presence within the genome data. ✗/✗ indicates that the enzyme is absent from both the genome and transcriptome data.

		AROMATIC AMINO ACIDS			
Steps	**ENZYMES**	*P. capuana*	*P. magna*	*P. parva*	*P. piriformis*
1	**SHKA1**	✓	✓	✓	✓
2	**DHQS**	✓	✓	✓	✓
3	**SHKD1**	✓	✓	✓	✓
4	**SHKF1**	✓	✓	✓	✓
5	**SHKG1**	✓	✓	✓	✓
6	**SHKH1**	✓	✓	✓	✓
7	**CHM1**	✓	✓	✓	✓
8	**AST1**	✓	✓	✓	✓
	AST2	✓	✓	✓	✓
	AST3	✓	✓	✓	✓
	AST4	✓	✓	✓	✓
	AST5	✗/✓	✓	✓	✓
	PAT	✓	✓	✓	✓
9a	**PRD1**	✓	✓	✓	✓
9b	**AGD1**	✓	✓	✓	✓
9c	**AAH1**	✓	✗/✓	✗/✓	✓
10a	**ANS1**	✓	✓	✓	✓
10b	**ASB1**	✓	✓	✓	✓
11	**PRT1**	✓	✓	✓	✓
12	**ASB2**	✓	✓	✓	✓
13	**IGS1**	✗/✗	✗/✗	✗/✗	✗/✗
14	**TSA**	✓	✓	✓	✓
15	**MAA7**	✓	✓	✓	✓

Enzymes leading to the direct synthesis of asparagine from aspartate, AST (isoforms 1-5) and ASNS, are present within all *Polytomella* species. AST5 is missing at the transcriptome but present at the genome level. All other enzymes involved in methionine and threonine biosynthesis were detected in the four species (see Table 8 below).

All of the enzymes involved in lysine biosynthesis were detected in the transcriptomes of *P. capuana*, *P. magna*, and *P. piriformis*. As shown in Table 8 below, *P. parva* is missing dihydrodipicolinate reductase (DPR) at the transcriptome level, but transcripts for the other lysine biosynthetic enzymes were identified. DPR is missing from the genome data of *P. piriformis*, despite its presence within the transcriptome. DPR is responsible for L-

2,3-dihydrodipicolinate conversion to tetrahydrodipicolinate, in the branch leading to lysine biosynthesis.

Briefly, enzymes required for the synthesis of methionine, threonine, asparagine, and lysine are all present within the four *Polytomella* species, with the exception of DPR.

Table 8: Aspartate-derived amino acid enzymes, where ✓ or ✗ represents the presence or absence of enzymes, respectively. ✗/✓ denotes the absence of the enzyme at the transcriptome level, but its presence within the genome data whereas ✓/✗ indicates that the enzyme is present within the transcriptome, but absent within the genome.

	ASPARAGINE, THREONINE, METHIONINE, AND LYSINE				
Steps	**ENZYMES**	*P. capuana*	*P. magna*	*P. parva*	*P. piriformis*
1	**AST1**	✓	✓	✓	✓
	AST2	✓	✓	✓	✓
	AST3	✓	✓	✓	✓
	AST4	✓	✓	✓	✓
	AST5	✗/✓	✓	✓	✓
2	**ASNS**	✓	✓	✓	✓
3	**ASK1**	✓	✓	✓	✓
4	**ASSD**	✓	✓	✓	✓
5	**HSD1**	✓	✓	✓	✓
	AHD1	✓	✓	✓	✓
6	**HSK1**	✓	✓	✓	✓
7	**THS1**	✓	✓	✓	✓
8	**CGS1**	✓	✓	✓	✓
9	**METC**	✓	✓	✓	✓
10	**METE**	✓	✓	✓	✓
	METH1	✓	✓	✓	✓
	METH2	✓	✓	✓	✓
11	**DPS1**	✓	✓	✓	✓
12	**DPR1**	✓	✓	✗/✓	✓/✗
13	**DPA1**	✓	✓	✓	✓
14	**DAE1**	✓	✓	✓	✓
15	**DPD1**	✓	✓	✓	✓

Analyses of *Polytomella* spp. genome and transcriptome data revealed the presence of all of the enzymes involved in the biosynthesis of branched chain amino, with the exception of branched-chain aminotransferase 3 (BCA3). However, other branched chain aminotransferase isoforms are present (Table 9).

Table 9: Branched-chain amino acid biosynthetic enzymes, where ✓ or ✗ represents the presence or absence of enzymes, respectively. ✗/✗ indicates that the enzyme is absent from both the genome and transcriptome data.

ISOLEUCINE, VALINE, AND LEUCINE					
Steps	**ENZYMES**	*P. capuana*	*P. magna*	*P. parva*	*P. piriformis*
1	**THD1**	✓	✓	✓	✓
2	**ALSL1**	✓	✓	✓	✓
	ALSS1	✓	✓	✓	✓
3	**AAI1**	✓	✓	✓	✓
4	**AAD1**	✓	✓	✓	✓
5	**BCA1**	✓	✓	✓	✓
	BCA2	✓	✓	✓	✓
	BCA3	✗/✗	✗/✗	✗/✗	✗/✗
6	**LEU2**	✓	✓	✓	✓
	LEU1L	✓	✓	✓	✓
	LEU1S	✓	✓	✓	✓
	LEU3a	✓	✓	✓	✓
	LEU3b	✓	✓	✓	✓

Finally, enzymes involved in arginine and proline biosynthetic pathways are present in all four species. The only missing enzyme is N-acetylornithine deacetylase (AOD or NAOD) from the transcriptomes and genomes of the four *Polytomella* species (table 10 below). This enzyme is responsible for ornithine formation from N-acetylornithine.

Table 10: Arginine and proline biosynthetic pathway, where ✓ or ✗ represents the presence or absence of enzymes, respectively. ✗/✗ indicates that the enzyme is absent from both the genome and transcriptome data.

ARGININE AND PROLINE				
ENZYMES	*P. capuana*	*P. magna*	*P. parva*	*P. piriformis*
AGK1	✓	✓	✓	✓
AGS1	✓	✓	✓	✓
AOD1	✗/✗	✗/✗	✗/✗	✗/✗
ARG1	✓	✓	✓	✓
ARG7	✓	✓	✓	✓
ARG9	✓	✓	✓	✓
CMPL1	✓	✓	✓	✓
CMPS1	✓	✓	✓	✓
GSD1/ P5CS	✓	✓	✓	✓
LCI8/ NAGS	✓	✓	✓	✓
OTA1/OAT	✓	✓	✓	✓
OTC1	✓	✓	✓	✓

PCR1	✓	✓	✓	✓
PROB1	✓	✓	✓	✓
PROB2	✓	✓	✓	✓

4.5 Discussion

The genetic and transcriptomic data indicate that *Polytomella* algae have an active and complete amino acid biosynthetic pathway and one that mirrors that of *C. reinhardtii*. Complete or near complete pathways for ammonia assimilation, branched chain, aromatic, glutamate-derived, and aspartate-derived amino acid biosynthesis were identified in all four species.

In a few cases, an enzyme was not identified in the transcriptome but found in the genome. For instance, GDH isoforms 1 and 2 and DPR in *P. parva*, and AAH1 in *P. magna* and *P. parva* were missing from transcriptome data but found within the genome. These results could be a consequence of an incomplete transcriptome due to Illumina sequencing errors, or redundant functionality in some cases and the presence of alternate pathways.

The majority of ammonia assimilation occurs through the GS/GOGAT cycle requiring two enzymes, glutamine synthase and glutamate synthetase. GDH has uncertain physiological roles within plants and green algae, but has been proposed to function in amino acid turnover (Fischer and Klein 1988; Cullimore and Sims 1981). The assimilation of ammonium and nitrate within land plants and green algae is the primary source of nitrogen required for subsequent dependent pathways such as amino acid biosynthesis. Nitrate (NO_3^-) is more often used due to its vast availability within soils and habitats as well as its regulatory roles within organisms (Scheible *et al.* 2004; Gerin *et al.* 2010). Nitrate uptake requires two transport and two reduction steps, while ammonia can permeate at high pH levels (basic conditions). Ammonium transporters are present in all of the four *Polytomella* species (can be accessed using GenBank accession numbers: AY542491/ AY542492). High-affinity nitrite (NO_2^-) transporters belonging to the formate class of transporters (NAR 1.6; XM_001696736 and AF149737) were also present in the *Polytomella* species and were more similar and hit to ones corresponding to *C. reinhardtii*, rather than other land plants and algae. In contrast, high-affinity nitrate

transporter of *C. reinhardtii* and *A. thaliana* (XM_001696376 and AF019748, respectively) were missing in the genome and transcriptome data of all *Polytomella* species. Nitrate reductase (XM_001696645), responsible for nitrate to nitrite conversion in the cytosol, and nitrite reductase (Y08937/ XM_001696375 and D14824), responsible for converting nitrite into ammonium, generated a no hit when blasted against the *Polytomella* transcriptome and genome data. The results imply that whereas ammonium and nitrite are being transported into the cell, nitrate reduction and transport into cells is missing. Similarly, nitrite reduction into ammonium does not seem to occur within the *Polytomella* plastids.

DPR, responsible for the reduction of L-2,3-dihydrodipicolinate into tetrahydrodipicolinate in the pathway branch leading to lysine, is missing from the transcriptome of *P. parva* and genome of *P. piriformis* but its presence at the genomic and transcriptomic levels respectively suggests that this enzyme is encoded and functional in each of these species. We even tried using different species DPR transcripts as query sequences with no luck in finding any homologous sequences in these two *Polytomella* species. One explanation could be the methodology by which Illumina sequencing has missed these transcripts, or the fact that *P. piriformis* may be synthesizing lysine using alternative pathways. Nonetheless, all other enzymes, including those for lysine, methionine, threonine, and asparagine are present within all four species. It is apparent that *Polytomella* follows *C. reinhardtii* and land plants in its non-abbreviated pathway to lysine biosynthesis. Often, in bacterial strains, an enzyme called meso-DAP dehydrogenase is responsible for the conversion of tetrahydrodipicolinate to lysine, eliminating the need for the four catalytic enzymes (Chatterjee *et al.* 1994; Misono *et al.* 1976). In plants and *C. reinhardtii*, meso-DAP dehydrogenase is absent, confirming the need for enzymes such as DPA, DAE, and DPD, and in fact, these enzymes were present. The presence of these enzymes within the four *Polytomella* species confirms its green algal origins and convergent evolution in terms of amino acid biosynthetic pathways with those of land plants and green photosynthetic algae.

AAH is responsible for an alternative pathway by which tyrosine is produced from phenylalanine. Its absence from the transcriptomes of *P. magna* and *P. parva*, despite its

detection within the genome data, may suggest that this alternative pathway has been omitted from the two species, leaving tyrosine to be synthesized only through arogenate dehydrogenase (AGD). This does not exclude the possibility of transcript loss or its inability to be detected using standard Illumina sequencing, and calls for further experimental studies to conclude its absence or presence accordingly.

Some enzymes, on the other hand, were missing entirely from genome and/or transcriptome data in the four *Polytomella* species. For instance, BCA isoform 3 of the branched chain amino acid biosynthesis, AOD of arginine and proline, and IGPS of aromatic amino acid biosynthesis were missing consistently of all *Polytomella* species.

Similar to land plants, *Polytomella* does not seem to use the linear pathway to ornithine biosynthesis as it is missing AOD. Upon failure to recover any transcript or gene sequences from any of the species using *C. reinhardtii* enzyme sequences as query sequences, we tried using *A. thaliana* enzyme sequence (AF360346), but still ended up with no success. It would be interesting to determine how *Polytomella* species synthesize arginine and whether AOD is truly missing *in vivo* or only missing from our transcriptome/genome sequencing data. Some plants share common enzymes with uretic organisms, using enzymes such as N-acetylornithine-glutamate aminotransferase (NAOGAT) and NAGS to make ornithine rather than relying on AOD (Slocum 2005). These enzymes are present in *Polytomella* suggesting that these species mimick land plants in their arginine biosynthesis. Moreover, there is also the possibility of arginine being imported into *Polytomella* cells (Kirk and Kirk 1978 a/b). More studies are needed, given that even within closely related plant species, AOD homology is low and therefore, sequence divergence may have influenced our lack of finding AOD in *Polytomella* spp (Slocum 2005).

Indole-3-glycerol phosphate synthase (IGPS or IGS) is missing from the transcriptome and genome sequence data of all four *Polytomella* species. IGS is crucial for the introduction of an indole ring for the formation of indole, and represents a branch point for the formation of tryptophan and indole-3-acetic acid (IAA), a major phytohormone in land plants (Ouyang *et al.* 2000; Östin *et al.* 1998). IAA is responsible for cell division,

signaling, root initiation, and senescence among other functions (Davies 1995; Östin *et al.* 1998). In addition to the absence of IGS in *Polytomella*, it has also been shown to be missing from *Helicosporidium* spp. (Pombert *et al.* 2014). It seems likely that our query sequences from photosynthetic land plants are inappropriate for use here; our species are nonphotosynthetic and unicellular, eliminating the need for many IAA physiological roles, and as such, the enzyme's presence within the species is questionable. Indeed, when using *Escherichia coli* IGS sequences as queries (GenBank accession number: AAC743443.3 and AAA57299.1), IGS was recovered in the transcriptomes of *P. capuana*, *P. parva*, and *P. piriformis*, and from the genome of *P. magna*. All corresponding hits from the NCBI database corresponded to other microalgae, green algae, and phytoplankton indole-3-glycerol phosphate synthases. It may be the case that *Polytomella*, along with other microalgae, have retained their ancestral versions of IGPS rather than having adopted or converged upon a plant version, which it does not require given that it is heterotrophic and small in size.

In summary, the amino acid biosynthetic pathways investigated all seem to occur within the *Polytomella* plastids. In some cases, a few enzymes were absent from either genome or transcriptome data, or both. The majority of these enzymes were found to be redundant in function, unneeded in heterotrophic and possibly free-living algae, or as of yet, undetermined function/significance in the pathways' progress.

4.6 Abbreviations

Ammonia assimilation and glutamate/glutamine biosynthesis enzymes

1. GDH1: glutamate dehydrogenase
   GDH2: glutamate dehydrogenase
2. GLN 1/2/3/4: glutamine synthetase
3. GSF1: Ferrodoxin-dependent glutamate synthase (Fd-GOGAT)
   GSN1: NADH-dependent glutamate synthase

Aromatic amino acid biosynthesis enzymes

1. SHKA1: 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (DAHP synthase)
2. DHQS: 3-dehydroquinate synthase (DHQ synthase)

3. SHKD1: bifunctional dehydroquinate dehydratase-shikimate: NADP oxidoreductase (DHQase-SORase)
4. SHKF1: shikimate kinase
5. SHKG1: 5-enolpyruvulshikimate-3-phosphate (EPSP) synthase
6. SHKH1: chorismate synthase (CS)
7. CHM1: chorismate mutase (CM)
8. PAT: prephenate aminotransferase
9. AST 1-5: Aspartate aminotransferase (AspAT)
10. ADT1: arogenate dehydratase
11. AGD1: arogenate/prephenate dehydrogenase
12. AAH1: aromatic amino acid hydroxylase
13. ANS1: anthranilate synthase, α subunit (ASα)
14. ASB1: anthranilate synthase, β subunit (ASβ)
15. PRT1: anthranilate phosphoribosyltransferase (APRT)
16. ASB2: phosphoribosylanthranilate isomerase (PAI)
17. IGS1: indole-3-glycerol-phosphate synthase (MAA4) (IGPS)
18. TSA: tryptophan synthetase, α-subunit (MAA1) (TSA)
19. MAA7: tryptophan synthase, β-subunit (TSB)

Aspartate-derived amino acid biosynthetic enzymes

1. AST 1-5: Aspartate aminotransferase (AspAT)
2. ASNS: asparagine synthetase (AS)
3. AHD1: bifunctional aspartate kinase/homoserine dehydrogenase (AK/HSDH)
   ASK1: aspartate kinase, monofunctional (AK)
4. ASSD: aspartate semialdehyde dehydrogenase (ASADH)
5. HSD1: homoserine dehydrogenase (HSDH)
6. HSK1: homoserine kinase (HSK)
7. THS: threonine synthase (TS)
8. CGS1: cystathionine gamma-synthase (CGS)
9. METC: cystathionine b-lyase (CBL)
10. METE: Cobalamin-independent methionine synthase (MS)
11. METH1/2: Cobalamin-dependent methionine synthase (MS)

12. DPS1: dihydrodipicolinate synthase (DHDPS)
13. DPR1: dihydrodipicolinate reductase (DHDPR)
14. DPA1: LL-diaminopimelate aminotransferase (DAP aminotransferase)
15. DAE1: diaminopimelate epimerase (DAPE)
16. DPD1: diaminopimelate decarboxylase (DAPD)

Branched chain amino acid biosynthetic enzymes

1. THD1: threonine deaminase (TD)
2. ALSL1: acetolactate synthase, large subunit (AHAS; ALS)
   ALSS1: acetolactate synthase, small subunit (AHAS; ALS)
3. AAI1: acetohydroxy acid isomeroreductase (AHRI or KARI)
4. AAD1: dihydroxyacid dehydratase (DHAD)
5. BCA1/2/3: branched-chain amino acid aminotransferase (BCAT)
6. LEU2: 2-isopropylmalate synthase
7. LEU1L: 3-isopropylmalate dehydratase, large subunit
   LEU1S: isopropylmalate dehydratase, small subunit
8. LEU3: 3-isopropylmalate dehydrogenase, splicing variants a and b

Arginine and proline biosynthetic enzymes

1. LCI8: N-acetylglutamate synthase (monofunctional) (NAGS)
2. AGK1: N-acetyl-L-glutamate kinase (NAGK)
3. ARG1: N-acetyl-γ-glutamyl-phosphate reductase (NAGPR)
4. ARG9: N-acetylornithine aminotransferase (NAOAT)
5. OTA1: ornithine-δ-aminotransferase (OAT)
6. AOD1: N-acetylornithine deacetylase (NAOD)
7. CMPL1: carbamoyl phosphate synthase, large subunit (CPS)
8. CMPS1: Carbamoyl phosphate synthase, small subunit (CPS)
9. OTC1: ornithine carbamoyltransferase
10. AGS1: Argininosuccinate synthase (AS)
11. ARG7: Argininosuccinate lyase (ASL)
12. GSD1: Δ^1- pyrroline-5-carboxylate synthetase (P5CS)
    PROB1/2: γ-glutamyl kinase

13. PCR1: Δ^1- pyrroline-5-carboxylate reductase (P5CR)

4.7 References

Akman, G., and Douglas, A.E. 2009. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. Proc Biol Sci 278; 987-991.

Alonso, E., and Rubio, V. 1989. Participation of ornithine aminotransferase in the synthesis and catabolism of ornithine in mice, studies using gabaculine and arginine deprivation. Biochem J 259: 131-138.

Chatterjee, S., Singh, B.K., and Gilvarg, C. 1994. Biosynthesis of lysine in plants: the putative role of meso-diaminopimelate dehydrogenase. Plant Molecular Biology 26: 285-290.

Coggins, J.R., Abell, C., Evans, L.B., Frederickson, M., Robinson, D.A., Roszak, A.W., and Lapthorn, A.P. 2003. Experiences with the shikimate-pathway enzymes as targets for rational drug design. Biochem Soc Trans 31: 548-552.

Crawford, N.M. 1995. Nitrate: nutrient and signal for plant growth. Plant Cell 7: 859-868.

Cullimore, J.V., and Sims, A.P. 1981. Occurrence of two forms of glutamate synthase in *Chlamydomonas reinhardtii*. Phytochemistry 20: 597-600.

Davies, P.J. 1995. The plant hormones: their nature, occurrence, and functions. In Plant Hormones (Davies, P.J., ed.). Dordrecht: Kluwer Academic Press, pp. 1-12.

Dewick, P.M. 1998. The biosynthesis of shikimate metabolites. Nat Prod Rep 15: 17-58.

Diebold, R., Schuster, J., Däschner, K., and Binder, S. 2002. The branched chain amino acid transaminase gene family in Arabidopsis encodes plastid and mitochondrial proteins. Plant Physiol 129: 540-550.

Fabro, G., Kovacs, I., Pavet, V., Szabados, L., and Alvarez, M.E. 2004. Proline accumulation and AtP5CS2 gene activation are induced by plant pathogen incompatible interactions in *Arabidopsis*. Mol Plant Microbe Interact 17: 343-350.

Facchini, P.J. 2001. Alkaloid biosynthesis in plants: biochemistry, cell biology, molecular regulation, and metabolic engineering application. Annu Rev Plant Physiol Plant Mol Biol 52: 29-66.

Ferreira, R.R., Vendemiatti, A., Gratão, P.L., Lea, P.J., and Azevedo, R.A. 2005. Determination of aspartate kinase in maize tissues. Sci Agric 62: 184-189.

Fischer, P., and Klein, U. 1988. Localization of nitrogen-assimilating enzymes in the chloroplast of *Chlamydomonas reinhardtii*. Plant Physiol 88: 947-952.

Grandoni, J.A., Marta, P.T., and Schloss, J.V. 1998. Inhibitors of branched-chain amino acid biosynthesis as potential antituberculosis agents. J Antimicrob Chemother 42: 475-482.

Gerin, S., Mathy, G., Blomme, A., Franck, F., and Sluse, F.E. 2010. Plasticity of the mitoproteome to nitrogen sources (nitrate and ammonium) in *Chlamydomonas reinhardtii*: the logic of Aox1 gene localization. Biochim Biophys Acta 1797: 994-1003.

Herrmann, K.M., and Weaver, L.M. 1999. The shikimate pathway. Annu Rev Plant Physiol Plant Mol Biol 50: 473-503.

Hoff, T., Truong, H.N., and Caboche, M. 1994. The use of mutants and transgenic plants to study nitrate assimilation. Plant, Cell Environ 17: 489-506.

Ikeda, M. 2006. Towards bacterial strains overproducing L-tryptophan and other aromatics by metabolic engineering. Appl Microbiol Biotechnol 69: 615-626.

Kirk, D.L., and Kirk, M.M. 1978a. Carrier-mediated uptake of arginine and urea by *Chlamydomonas reinhardtii*. Plant Physiol 61: 556-560.

Kirk, M.M., and Kirk, D.L. 1978b. Carrier-mediated uptake of arginine and urea by *Volvox carteri* f. nagariensis. Plant Physiol 61: 549-555.

Kropat, J., Hong-Hermesdorf, A., Casero, D., Ent, P., Castruita, M., Pellegrini, M., Merchant, S.S., and Malasarn, D. 2011. A revised mineral nutrient supplement increases biomass and growth rate in *Chlamydomonas reinhardtii*. Plant J 66: 770-780.

Knaggs, A.R. 2001. The biosynthesis of shikimate metabolites. Nat Prod Rep 18: 334-355.

Knaggs, A.R. 2003. The biosynthesis of skikimate metabolites. Nat Prod Rep 20: 119-136.

Less, H., and Galili, G. 2008. Principal transcriptional programs regulating plant amino acid metabolism in response to abiotic stresses. Plant Physiol 147: 316-330.

Layne, E. 1957. Spectrophotometric and turbidimetric methods for measuring proteins. Methods Enzymol 3: 447-455.

Leuchtenberger, W., Huthmacher, K., and Drauz, K. 2005. Biotechnological production of amino acids and derivatives: current status and prospects. Appl Microbiol Biotechnol 69: 1-8.

Marchler-Bauer, A., Anderson, J.B., DeWeese-Scott, C., Fedorova, N.D., Geer, L.Y., He, S., Hurwitz, D.I., et al. 2003. CDD: a curated Entrez database of conserved domain

alignments. Nucl. Acids Res. 31: 383–387 (CDD v. 2.02; http://www.ncbi.nlm.nih.gov/ Structure/cdd/cdd.shtml).

Miflin, B.J., and Lea, P. 1976. The pathway of nitrogen assimilation in plants. Phytochemistry 15: 873-885.

Misono, H., Togawa, H., Yamamoto, T., Soda, K. 1979. m-α-ε-Diaminopimelate dehydrogenase: distribution and the reaction product. J Bact 137: 22-27.

Mueller, U., and Huebner, S. 2003. Economic aspects of amino acids production. In T Scheper, ed., Advances in Biochemical Engineering/Biotechnology, Springer-Verlag, Berlin, 79: 137-170.

Nichols, B.P. 1996. Evolution of genes and enzymes of tryptophan biosynthesis. In: Neidhart FC (ed) *Escherichia coli* and *Salmonella*: cellular and molecular biology, vol. 2. ASM, Washington, DC, pp. 2638-2648.

Östin, A., Kowalyczk, M., Bhalerao, R.P., and Sandberg, G. 1998. Metabolism of indole-3-acetic acid in *Arabidopsis*. Plant Physiol 118: 285-296.

Ouyang, J., Shao, X., and Li, J. 2000. Indole-3-glycerol phosphate, a branchpont of indole-3-acetic acid biosynthesis from the tryptophan biosynthetic pathway in *Arabidopsis thaliana*. Plant J 24: 327-333.

Pátek, M. 2007. Branched-chain amino acids. Microbiol Monogr 5: 129-162.

Pittard, J. 1996. Biosynthesis of the aromatic amino acids. In: Evolution of genes and enzymes of tryptophan biosynthesis. In: Neidhart FC (ed) *Escherichia coli* and *Salmonella*: cellular and molecular biology, vol. 1. ASM, Washington, DC, pp. 458-484.

Ralph, S.A., van Dooren, G.G., Waller, R.F., Crawford, M.J., Fraunholz, M.J., Foth, B.J., Tonkin, C.J., Roos, D.S., and McFadden, G.I. 2004. Tropical infectious diseases: metabolic maps and functions of the *Plasmodium falciparum* apicoplast. Nat Rev Microbiol 2: 203-216.

Rosenthal, G.A. 1982. L-Canavanine metabolism in Jack Bean, *Canavalia ensiformis* (L.) DC. (Leguminosae). Plant Physiol 69: 1066-1069.

Scheible, W.R>, Morcuende, R., Czechowski, T., Fritz, C., Osuna, D., Palacios-Rojas, N., Schindelasch, D., Thimm, O., Udvardi, M.K., and Stitt, M. 2004. Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes, and the regulatory infrastructure of *Arabidopsis* in response to nitrogen. Plant Physiol 136: 2483-2499.

Schuster, J., and Binder, S. 2005. The mitochondrial branched chain aminotransferase (AtBCAT-1) is capable to initiate degradation of leucine, isoleucine, and valine in almost all tissues in *Arabidopsis thaliana*. Plant Mol Biol 57: 241-254.

Shargool, P.D., Jain, J.C., and McKay, G. 1988. Ornithine biosynthesis and arginine biosynthesis and degradation in plant cells. Phytochemistry 27: 1571-1574.

Sharma, S., and Verslues, P.E. 2010. Mechanisms independent of ABA or proline feedback have a predominant role in transcriptional regulation of proline metabolism during low water potential and stress recovery. Plant Cell Environ 33: 1838-1851.

Singh, B.K. 1999. Biosynthesis of valine, leucine and isoleucine. In BK Singh, ed., Plant amino acids: Biochemistry and biotechnology. Marcel Dekker, New York, pp 227-247.

Slocum, R.D. 2005. Genes, enzymes and regulation of arginine biosynthesis in plants. Plant Physiol Biochem 43: 729-745.

Srinivasan, P.R., and Sprinson, D.B. 1959. 2-keto-3-deoxy-D-arabo-hepatonic acid 7-phosphate synthetase. J Biol Chem 234: 713-716.

Szekely, G., Abraham, E., Cselo, A., Rigo, G., Zsigmond, L., Csiszar, J., Ayaydin, F., Strizhov, N., Jasik, J., Schmeizer, E., Koncz, C., and Szabados, L. 2008. Duplicated P5CS genes of *Arabidopsis* play distinct roles in stress regulation and developmental control of proline biosynthesis. Plant J 53: 11-28.

Thompson, J.F. 1980. Arginine synthesis, proline synthesis, and related processes. In Miflin, B.J. (Ed.) The biochemistry of plants, vol. 5. Academic Press Inc., New York, pp. 375-402.

Verslues, P.E., and Bray, E.A. 2006. Role of abscisic acid (ABA) and *Arabidopsis thaliana* ABA-insensitive loci in low water potential-induced ABA and proline accumulation. J Exp Bot 57: 201-212.

Wang, J.G., Li, Z.M., Ma, N., Wang, B.L., Jiang, L., Pang, S.S., Lee, Y.T., Guddat, L.W., Duggleby, R.G. 2005. Structure-activity relationships for a new family of sulfonylurea herbicides. J Comput Aided Mol Des 19: 801-820.

Yanofsky, C., and Crawford, I.P. 1972. Tryptophan synthetase. In: Boyer PD (ed.) The Enzymes, vol. 7, 3rd ed. Academic, New York, pp. 1-31.

Yokota, A., Kawasaki, S., Iwano, M., Nakamura, C., Miyake, C., and Akashi, K. 2002. Citrulline and DRIP-1 protein (ArgE homologue) in drought tolerance of wild watermelon. Ann Bot 89: 825-832.

Chapter 5

A condensed and slightly different version of this chapter, with alternative figures, is currently under review at the journal *New Phytologist* (Asmail S and Smith DR. Retention, erosion, and loss of the carotenoid biosynthetic pathway in the nonphotosynthetic green algal genus *Polytomella*. Submitted 30 June 2015). A complete version of this submission including tables, figures, and supplementary materials is found in Appendix B.

5 The Carotenoid Biosynthetic Pathway

5.1 Introduction: carotenoids at a glance.

Over 600 organic pigments, known as carotenoids, are synthesized within the chloroplasts of photosynthetic organisms, and in some nonphotosynthetic species, including certain bacteria and fungi (Straub 1987; Kull and Pfander 1995). Carotenoids are isoprenoid molecules, composed of a polyene chain of conjugated bonds that are symmetrical around the central double bond (Britton 1995). The polyene chain acts as a chromophore, determining the light absorption spectra and thus giving the carotenoids their characteristic colors (Britton 1995). Moreover, chromophores also determine the photochemical properties of carotenoids, making them responsible for beneficial antioxidant effects (Britton 1995). There are two general carotenoid structures: carotenes and xanthophylls. Polyene chains that are cyclized at one or both ends are known as carotenes whereas xanthophylls are the oxidized versions of the polyene chain, and both are located along with chlorophyll and other light-harvesting complexes in the thylakoid membrane within the plastid (Demmig-Adams 1990; Niyogi 1999).

The functions of carotenoids are derived from their inherent structure. The conjugated polyene chain, formed of alternating double and single bonds, gives carotenoids diversity as each bond could form an isomer, leading to a different carotenoid molecule (Britton 1995). This diversity in structure leads to diversity in function, ranging from stabilizing reactive oxygen species (or ROS), to formation of retinal, to helping with membrane structure and rigidity, to protecting against the most serious of human diseases like cancer

(Britton 1995; Lazrak *et al.* 1987; Krinsky 1989; Bendich 1989; Zhang *et al.* 1991). Carotenoids and xanthophylls within the plastids (AKA chromoplasts) of photosynthetic plants absorb light at different wavelengths than chlorophyll (450-570 nm), acting as accessory light-harvesting pigments, and produce a variety of vibrant colors like red, yellow, and green to attract predators and pollinators alike, helping the organism to survive and thrive (Britton 1995; Dieckmann 2003; Sineshchekov *et al.* 2002). Despite their role in harvesting photosynthetic energy and thylakoid membrane formation, they also protect the organism against excessive light and related photo-oxidative damage (Herrin *et al.* 1992). In plants and green algae, carotenoids have proved useful in combatting abiotic stresses. In *Chlamydomonas reinhardtii*, carotenoids have been shown to mediate the faster recovery of damaged PSII complexes, increasing their tolerance to oxidative stress (Pardha-Saradhi *et al.* 2000, Yoshida *et al.* 2003, 2004). Carotenoids have also been hypothesized to result in cyst formation in drought-stressed *C. reinhardtii* zygospores (Kobayashi *et al.* 1997).

In unicellular flagellated organisms, such as *C. reinhardtii* and *Plasmodium falciparum*, carotenoids accumulate in a specific phototactic organelle, called the eyespot (sometimes known as stigma). First studied in the late 1800s, the eyespot is a photosensitive structure, responsible for not only absorbing different wavelengths of light, but also for protecting the organism against excessive light (Faminzin 1867; Mast 1917; Kröger and Hegemann 1994; Herrin *et al.* 1992). The eyespot is commonly found at the anterior end within *C. reinhardtii* resulting in asymmetrical cells, and consists of thylakoid membrane and plasma membrane constituents (Britton 1995). The carotenoid granules direct light towards a retinal-binding rhodopsin homologue photoreceptor, allowing the cells to swim towards or away from the light source (Kröger and Hegemann 1994). In eyespot-lacking mutants, the cells are still able to respond to light via the rhodopsin photoreceptor present in the plasma membrane but with less speed and direction (Hartshorne 1953; Harris 2001; Dieckmann 2003; Foster *et al.* 1984). Directing cells away from extreme light conditions requires flagellar membrane depolarization via a calcium-dependent all or none electric current (Ehlenbeck *et al.* 2002; Harz and Hegemann 1991; Dieckmann 2003). In 1917, Buder studied *Polytomella* and *Polytoma* to find that only species with the eyespot were

sensitive to sunlight. Carotenoid biosynthesis was then left unstudied in these latter species, until now.

Recent analyses have characterized multiple enzymes culminating in 15 steps responsible for carotenoid biosynthesis in *C. reinhardtii*. This was done mostly through comparative genomics and various HPLC and TLC techniques that have been developed for photosynthetic land plants and green algae (Armstrong and Hearst 1996; Cunningham and Gantt 1998; Hirschberg 2001; Bouvier *et al.* 2005; Grossman *et al.* 2004; Lohr *et al.* 2005). Carotenoid biosynthesis occurs in the plastids of green algae and land plants, with the enzymes being encoded in the nuclear genome, synthesized in the cytosol, and then imported via a plastid-targeting sequence into the plastid/chloroplast (Sirevåg and Levine 1973, Lohr *et al.* 2005). The process of carotenoid synthesis consists of various reactions starting with condensation and reduction reactions to make the first 40-carbon carotenoid, phytoene (Schwender *et al.* 1999; Lohr *et al.* 2005; Sun *et al.* 1998; McCarthy *et al.* 2004). Then desaturation and isomerization reactions result in the red carotenoid, lycopene (McCarthy *et al.* 2004; Nikulina *et al.* 1999). Afterwards, a series of cyclization and hydroxylation reactions occur to end up with various other carotenes and xanthophylls, respectively (Lohr *et al.* 2005; Anwaruzzaman *et al.* 2004; Baroli *et al.* 2003).

5.2 Carotenoids in detail.

The carotenoid biosynthetic pathway, described in detail in Figure 12, starts with the condensation of glyceraldehyde-3-phosphate with thiamine (obtained from pyruvate) to result in 1-deoxy-D-xylulose-5-phosphate (DXP). DXP is then reduced to 2-C-methyl-D-erythritol-4-phosphate (or MEP) which is converted to 2-C-methyl-D-erythritol-2,4-cyclodiphosphate (ME-cPP) with the help of 3 enzymes namely; CMS (4-diphosphocytidyl-2-C-methyl-D-erythritol synthase), CMK (4-diphosphocytidyl-2-C-methyl-D-erythritol kinase), and MCS (ME-cPP synthase) respectively (Figure 12). ME-cPP becomes 4-hydroxy-3-methylbut-2-enyldiphosphate (HMBPP), under the influence of HMBPP synthase (HDS) and results in the production of a mixture of 5 IPP (isopentenyl-diphosphate) to 1 DMAPP (dimethylallyl-diphosphate) which can interconvert with the help of IPP isomerase (IDI). The combination of 3 IPP to 1 DMAPP

results in a 20-carbon molecule geranylgeranyl diphosphate (GGPP). The condensation of 2 GGPP molecules results in a 40-carbon molecule and the first carotenoid, phytoene, with the help of geranylgeranyl diphosphate synthase (GGPPS) and phytoene synthase (PSY), respectively. The following steps lead to the red pigment lycopene, found in red plants and fruits such as tomatoes, and involve a series of desaturations increasing the number of double bonds as well as isomerizations. Phytoene is converted to phytofluene (with 5 double bonds) via phytoene desaturase (PDS). The desaturation step is once again repeated, resulting in 7 double bonds and the carotenoid that is Z-carotene, a reaction catalyzed by Z-carotene isomerase (Z-ISO). Z-carotene is desaturated to make neurosporene with 9 double bonds in the polyene chain, with a characteristic orange/yellow color, via the help of Z-carotene desaturase (ZDS). The polyene chain is further reduced to end up with 11 double bonds and a pink/red color, and resulting in lycopene via the activity of carotenoid isomerase (CRTISO). The cyclization of the polyene chain ends results in carotenes, with α-carotenes having 1β and 1 ε rings, whereas β-carotenes possess 2β rings, one at each end of the chain.

Lycopene-β-cyclase (LCYB) and lycopene-ε-cyclase (LCYE) are responsible for the addition of these rings to the ends of the polyene chain. The hydroxylation of these rings results in xanthophylls, the second major carotenoids category, some of which help with the dissipation of excessive sunlight energy in land plants. The reactions are catalyzed by either carotene-β-hydroxylase (CHYB)—responsible for lutein production from α-carotene, and carotene-ε-hydroxylase (CHYE)—responsible for zeaxanthin production from β-carotene. The epoxidation of zeaxanthin via zeaxanthin epoxidase (ZEP) results in violaxanthin, and has been shown to work particularly well in the dark. Another enzyme, violaxanthin deepoxidase, which results in the reverse reaction, has been shown to function in the light instead. Further modification of violaxanthin can result in neoxanthin, a reaction catalyzed by neoxanthin synthase (NSY). However, this step of the carotenoid biosynthetic pathway has not been shown to work in *C. reinhardtii*, although it is present in *A. thaliana* (Lohr 2009). Further steps result in the acylation of the carotenoids via the work of BKT, LUT1, and LUT5. It is still unknown the specific consequence of these acylation reactions, but are suggested to function in sequestering

carotenoids to lipid granules and expanding the carotenoids expiry date by preventing chromophore crystallization (Lohr 2009).

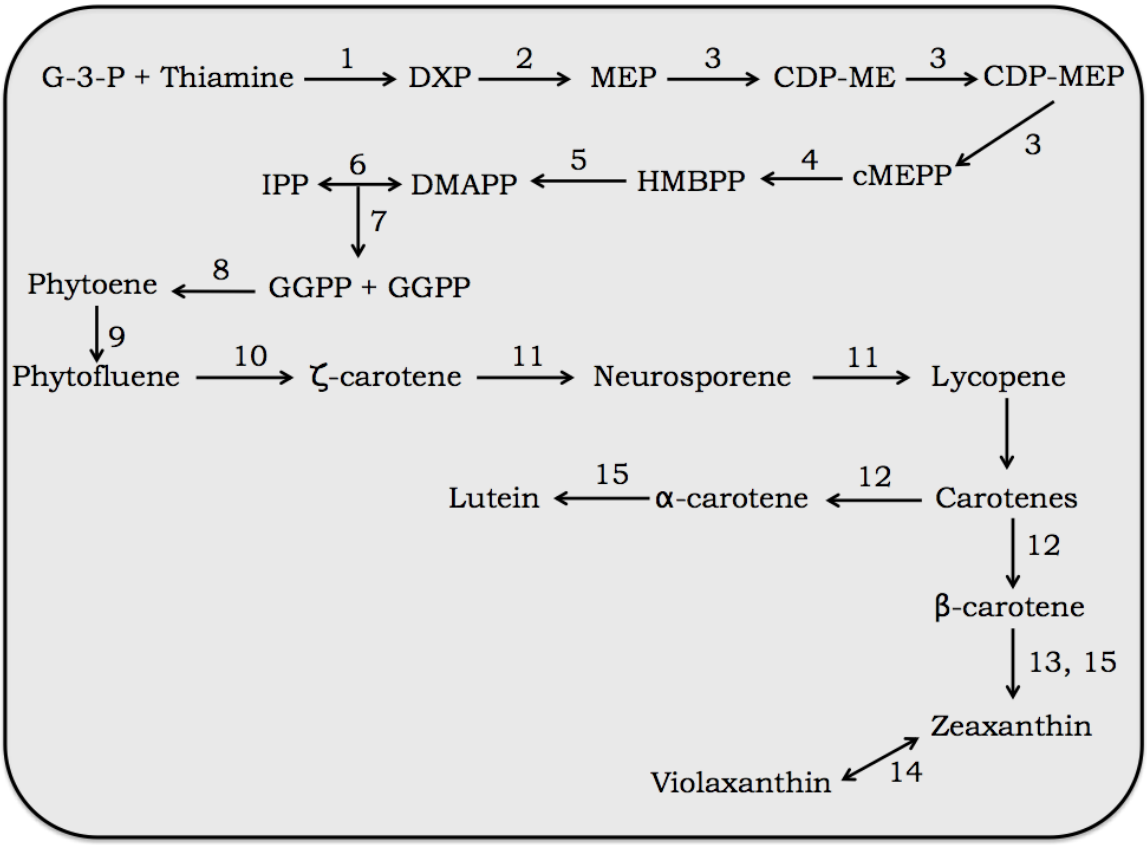


Figure 12: The carotenoid biosynthetic pathway. *G-3-P*, glyceraldehyde-3-phosphate; *DXP*, 1-deoxy-D-xylulose-5-phosphate; *MEP*, 2-C-methyl-D-erythritol-4-phosphate; *CDP-MEP*, 2-phospho-4-(cytidine 5′-diphospho)-2-C-methylD-erythritol; *cMEPP*, 2-C-methyl-D-erythritol-2,4-cyclodiphosphate; *HMBPP*, 4-hydroxy-3-methylbut-2-enyldiphosphate; *DMAPP*, dimethylallyl-diphosphate; *IPP*, isopentenyl-diphosphate; *GGPP*, geranylgeranyl diphosphate. Numbers correspond to the enzymes in table 11 below, in order of their appearance. 1, DXS or 1-deoxy-D-xylulose-5-phosphate synthase; 2, *DXR* or 1-deoxy-D-xylulose-5-phosphate reductoisomerase; 3, *CMS* or 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase, *CMK* or 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, *MCS* or 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase; 4, *HDS* or 4-hydroxy-3-methylbut-2-enyl-1-diphosphate synthase; 5, *IDS* or isopentenyl/dimethylallyl diphosphate synthase;

6, *IDI* or isopentenyl/dimethylallyl diphosphate Δ isomerase; 7, *GGPPS* or geranylgeranyl diphosphate synthase; 8, *PSY* or phytoene synthase; 9, *PDS* or phytoene desaturase; 10, *Z-ISO* or ζ-carotene isomerase; 11, *ZDS* or ζ-carotene desaturase, *CRTISO* or carotenoid isomerase; 12, *LCYB* or lycopene-β-cyclase, *LCYE* or lycopene-ε-cyclase; 13, *CHYB* or carotene-β-hydroxylase; 14, *ZEP* or zeaxanthin epoxidase, *VDE* or violaxanthin de-epoxidase; 15, *BKT* or carotene-β-ketolase, *LUT1* or carotene-ε-hydroxylase, *LUT5* or carotene-β-hydroxylase. Full enzyme names are provided in Section 5.6, and GenBank accession numbers are listed in the list of abbreviations.

5.3 Methods

Please refer to section 2.3 for detailed explanation on transcriptomic and genomic data analyses.

In summary, local databases for each of the four *Polytomella* species were created, and enzymes involved in the carotenoid biosynthetic pathway in *C. reinhardtii* and *A. thaliana* were used as query sequences to search these local databases. Resulting hits were once again checked against NCBI to further confirm that the resulting hits were in fact the enzymes of interest. For specific species that were missing particular enzymes, the *Polytomella* genome was under investigation, to see whether these enzymes were encoded but were not expressed, or whether they were missing entirely from the *Polytomella* transcriptome. Hits were once again checked with NCBI non-redundant database to further confirm the accuracy and validity of these hits.

5.4 Results

Our results suggest the progressive loss of the carotenoid pathway across the *Polytomella* species. In *P. magna*, the majority of the genes encoding carotenoid pathway enzymes were present. *P. magna* is the most basal species of the *Polytomella* genus, and is the least-derived in phylogenetic relation with *C. reinhardtii*. Carotenoid enzyme mRNA sequences of *C. reinhardtii*, with NSY (neoxanthin synthase) missing in the green algae, but present in *A. thaliana* were used as query sequences to search the *Polytomella*

database, in the hope of finding homologues within the *Polytomella* species. NSY was missing in all four *Polytomella* species. *P. magna* has also lost some of the final steps downstream of β-carotene, namely, violaxanthin de-epoxidase (VDE) and zeaxanthin epoxidase (ZEP). This was further confirmed by genome analyses, which showed the absence of genes for both enzymes. Moreover, in *P. magna*, geranylgeranyl reductase (GGR) is also missing from the transcriptome and genome data. LUT5, involved in β-ring hydroxylation of α- and β-carotenes—an essential step in xanthophyll production (Fiore et al. 2006), is also missing from both datasets.

Similarly, *P. capuana* was also missing all of the genes encoding the enzymes mentioned above, with the addition of BKT from its transcriptome despite its presence at the genome level.

On the more extreme, *P. parva* and *P. piriformis* fully lack the carotenoid biosynthetic pathway and have only retained the initial steps within the pathway. *P. piriformis* lacks all the genes encoding enzymes after geranylgeranyl diphospate synthase (GGPPS), at the transcriptome and genome levels.

5.5 Discussion

In *P. magna*, which has been shown to possess a pinkish eyespot (Pringsheim 1955) similar to that of *C. reinhardtii*, the majority of the carotenoid enzymes were present at the transcriptome and genome levels. GGR is missing at both levels, across all four species, demonstrating the efficacy by which the *Polytomella* lineages have gotten rid of any photosynthetic-related machinery. GGR is responsible for geranylgeranyl pyrophosphate reduction to produce phytyl diphosphate, one of the precursors for chlorophyll biosynthesis from chlorophyllide (Grossman *et al.* 2004). ZEP and VDE are responsible for nonphotochemical quenching and are part of the light-harvesting complexes in land plants and green algae (Grossman *et al.* 2004; Niyogi 1999). VDE catalyzes the reverse reaction, i.e. the de-epoxidation of violaxanthin that was formed via ZEP, and is responsible for the dissipation of excessive light energy in light-harvesting complexes in land plants and photosynthetic green algae (Niyogi 1999). Thus, their absence across all *Polytomella* lineages confirms the enzymes' functional role in

photosynthetic-related processes, and their dispensability in nonphotosynthetic species that have omitted such non-relevant enzymes from their genomes. The absence of LUT1 and LUT5 suggest that *Polytomella* spp. are unable to synthesize lutein (Tian *et al.* 2004). BKT is missing in *P. capuana* at the transcriptome level but is encoded within the genome. This suggests that it is not expressed in the transcriptome due to the fact that this species does not possess carotenoids and as such there is no need for a lipid-carotenoid association (Lohr 2009).

P. piriformis and *P. parva* are missing all of the enzymes committed to carotenogenesis. Carotenoid biosynthesis is not expected to occur within these species due to their lack of carotenoid-specific enzymes as shown by our results, and indeed, pigmentation and eye spot retention has not been observed in the two species (Pringsheim 1955). In fact, PSY mutants in *C. reinhardtii* show similar characteristics to that of *Polytomella*, with chloroplasts that are bleached and lacking any carotenoid pigments, with many large and variable-sized starch granules, and irregular eyespot conformation and location (Inwood *et al.* 2008).

There are varying explanations as to why there is a progressive loss within these organisms. One might be that *P. magna* has not lost its carotenoid pathway yet, merely for the fact that *P. magna* is the most closely related species to photosynthetic *C. reinhardtii* and may be not as derived as the other species. Another explanation as to why *P. magna* has not lost its carotenoid biosynthetic pathway or its eyespot is that this species may need it to navigate away or towards the sunlight. *Polytomella* is a nonphotosynthetic, heterotrophic genus that does not require photosynthesis for its survival. The presence of the carotenoid pathway in nonphotosynthetic plastid-containing organisms is rare, but not unprecedented. In *P. falciparum*, a parasitic malaria-causing agent with a plastid of red algal origins, the plastid is responsible for a variety of metabolic pathways, one of which is carotenoid biosynthesis (Borza *et al.* 2005; Ralph *et al.* 2004).

Table 11: Carotenoid biosynthesis enzymes, where ✓ or ✗ represents the presence or absence of enzymes, respectively. ✗/✗ indicates that the enzyme is absent from both the genome and transcriptome data. ✗/✓ denotes that absence of the enzyme from the *Polytomella* transcriptome, but its presence within the genome data.

Step	Carotenoid Enzymes	*C. reinhardtii*	*P. magna*	*P. capuana*	*P. piriformis*	*P. parva*
1.	DXS	✓	✓	✓	✓	✓
2.	DXR	✓	✓	✓	✓	✓
3.	CMS	✓	✓	✓	✓	✓
	CMK	✓	✓	✓	✓	✓
	MCS	✓	✓	✓	✓	✓
4.	HDS	✓	✓	✓	✓	✓
5.	IDS	✓	✓	✓	✓	✓
6.	IDI	✓	✓	✓	✓	✓
7.	GGPPS	✓	✓	✓	✓	✓
	GGR	✓	✗/✗	✗/✗	✗/✗	✗/✗
8.	PSY	✓	✓	✓	✗/✗	✗/✗
9.	PDS	✓	✓	✓	✗/✗	✗/✗
10.	Z-ISO	✓	✓	✓	✗/✗	✗/✗
11.	ZDS	✓	✓	✓	✗/✗	✗/✗
	CRTISO	✓	✓	✓	✗/✗	✗/✗
12.	LCYB	✓	✓	✓	✗/✗	✗/✗
	LCYE	✓	✗/✗	✗/✗	✗/✗	✗/✗
13.	CHYB	✓	✓	✓	✗/✗	✗/✗
14.	ZEP	✓	✗/✗	✗/✗	✗/✗	✗/✗
	VDE	✓	✗/✗	✗/✗	✗/✗	✗/✗
15.	BKT	✓	✓	✗/✓	✗/✗	✗/✗
	LUT1	✓	✗/✗	✗/✗	✗/✗	✗/✗
	LUT5	✓	✗/✗	✗/✗	✗/✗	✗/✗

A study by Tonhosolo *et al.* (2009) showed that not only do these apicomplexan parasites produce a variety of carotenoids, but that these carotenoids are in fact essential for parasite growth. Recent studies have suggested that targeting carotenogenesis and isoprenoid biosynthesis in these parasites could be of great medical use; since it would be targeted to the parasites rather than hosts (humans and/or other vertebrates) which do not possess this particular pathway (Coppens 2013; Tonhosolo *et al.* 2009; Jomaa *et al.* 1999; Wiesner *et al.* 2008; Hunter 2011; Lell *et al.* 2003; Rodrigues *et al.* 2004). Carotenoids are essential for protection against reactive oxygen molecules, acting as scavengers, and protecting lipophilic membranes against oxidation (Britton 1995; Lazrak *et al.* 1987; Krinsky 1989; Bendich 1989). It could be the result of an oxidative environment, that it has been retained in *P. magna*. That raises the question of why carotenoid biosynthesis has not been retained in the other *Polytomella* species explored here? It may be that these species either live in favorable environments (with no light and/or oxygen), that they have evolved some other ROS-scavenging technique, or lost it by chance and random mutation with time and have adapted to this loss.

Despite the mystery surrounding why carotenoid biosynthesis has been retained in one, and lost in the three other species, the study of carotenogenesis provides insight into the loss of photosynthesis in *Polytomella* and other related species. Several studies have suggested that thylakoid formation is fully dependent on the presence of carotenoids and that the origin of nonphotosynthetic species from a photosynthetic ancestor may lie with point mutations in phytoene synthase (Herrin *et al.* 1992; Rumpf *et al.* 1996; Inwood *et al.* 2008). We now know that is not the case, at least in *Polytomella*. The beauty of *Polytomella* lies in that, within the same phylogenetic group, we have multiple lineages displaying a progressive loss of carotenoid biosynthesis, with some retaining it fully, and others that have lost it but at various stages, genetically speaking. In *Polytomella*, the absence of thylakoid membranes is observed across all species, regardless of whether carotenoids are absent or present. Moreover, phytoene synthase may not be the key initiator of nonphotosynthetic species, as phytoene synthase is present in two of our nonphotosynthetic species, namely, *P. magna* and *P. capuana*, with only the former possessing carotenoids. In addition, we suggest that carotenoid loss has occurred twice

within the *Polytomella* genus, once in the *P. capuana* branch and the other in the phylogenetic branch leading to *P. parva*/ *P. piriformis* resulting in the similar loss of enzymes in the species.

It is the conclusion of this work that *Polytomella* would be a great model organism to study eyespot formation, development, and function in these nonphotosynthetic species, to possibly correlate their living and environmental conditions to carotenoid retention or loss, to investigate the genetic and environmental differences between *P. magna* and other *Polytomella* spp. that has made the former possess an eyespot and a fully pigmented cell while the latter not, and finally, to understand the mechanisms and adaptations that have made these carotenoid-less cells able to perform well in light and oxygenic environments.

5.6 Abbreviations

1. DXS: 1-deoxy-D-xylulose-5-phosphate synthase
2. DXR: 1-deoxy-D-xylulose-5-phosphate reductoisomerase
3. CMS: 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase
4. CMK: 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase
5. MCS: 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase
6. HDS: 4-hydroxy-3-methylbut-2-enyl-1-diphosphate synthase
7. IDS: isopentenyl-/dimethylallyl-diphosphate synthase
8. IDI: isopentenyl-diphosphate delta-isomerase
9. GGPPS: geranylgeranyl-diphosphate synthase
10. PSY: phytoene synthase
11. Z-ISO: zeta-carotene isomerase
12. ZDS: zeta-carotene desaturase
13. CRTISO: Carotenoid isomerase
14. LCYB: Lycopene-β-cyclase
15. LCYE: Lycopene-ε-cyclase
16. CHYB: Carotene-β-hydroxylase
17. ZEP: Zeaxanthin epoxidase
18. VDE: Violaxanthin de-epoxidase

19. NSY: Neoxanthin synthase
20. BKT: Carotene-β-ketolase

5.7 References

Anwaruzzaman, M., Chin, B.L., Li, X.P., Lohr, M., Martinez, D.A., and Niyogi, K.K. 2004. Genomic analysis of mutants affecting xanthophyll biosynthesis and regulation of photosynthetic light harvesting in *Chlamydomonas reinhardtii*. Photosynth Res 82: 265-276.

Armstrong, G.A., and Hearst, J.E. 1996. Carotenoids. 2. Genetics and Molecular biology of carotenoid pigment biosynthesis. FASEB J 10: 228-237.

Baroli, I., Do, A.D., Yamane, T., and Niyogi, K.K. 2003. Zeaxanthin accumulation in the absence of a functional xanthophyll cycle protects *Chlamydomonas reinhardtii* from photooxidative stress. Plant Cell 15: 992-1008.

Bendich, A. 1989. Carotenoids and the immune response. J Nutr 119: 112-115.

Bouvier, F., Rahier, A., and Camara, B. 2005. Biogenesis, molecular regulation and function of plant isoprenoids. Prog Lipid Res 44: 357-429.

Borza, T., Popescu, C.E., and Lee, R.W. 2005. Multiple metabolic roles for the nonphotosynthetic plastid of the green alga *Prototheca wickerhamii*. Eukaryo Cell 4: 253-261.

Buder, J. 1917. Zur Kenntnis der phototaktischen Richtungsbewegungen. Jb wiss Bot 58: 105.

Britton, G. 1995. Structure and protperties of carotenoids in relation to function. FASEB J 9: 1551-1558.

Coppens, I. 2013. Targeting lipid biosynthesis and salvage in apicomplexan parasites for improved chemotherapies. Nature Reviews 11: 823-835.

Cunningham, F.X., Jr and Gantt, E. 1998. Genes and enzymes of carotenoid biosynthesis in plants. Annu Rev Plant Physiol Plant Mol Biol 49: 557-583.

Demmig-Adams, B. 1990. Carotenoids and photoprotection in plants: a role for the xanthophyll zeaxanthin. Biochim Biophys Acta 1020: 1-24.

Dieckmann, C.L. 2003. Eyespot placement and assembly in the green alga *Chlamydomonas*. BioEssays 25: 410-416.

Ehlenbeck, S., Gradmann, D., Braun, F.J., Hegemann, P. 2002. Evidence for a light-induced H+ conductance in the eye of the green alga *Chlamydomonas reinhardtii*. Biophys J 82: 740-751.

Faminzin, A. 1867. Die Werkung des Lichtes auf algen und einige andere nahe verwandte Organismen. Jb wiss Bot 6: I.

Fiore, A., Dall'Osto, L., Fraser, P.D., Bassi, R., and Giuliano, G. 2006. Elucidation of the B-carotene hydroxylation pathway in *Arabidopsis thaliana*. FEBS Letters 580: 4718-4722.

Foster, K.W., Saranak, J., Patel, N., Zarilli, G., Okabe, M., Kline, T., Nakanishi, K. 1984. A rhodopsin is the functional photoreceptor for phototaxis in the unicellular eukaryote *Chlamydomonas*. Nature 311: 756-759.

Goulart, H.R., Kimura, E.A., Peres, V.J., Couto, A.S., Duarte, F.A.A., and Katzin, A.M. 2004. Terpenes arrest parasite development and inhibit biosynthesis of isoprenoids in *Plasmodium falciparum*. Antimicrobiol Agents Chemother 48: 2502-2509.

Grossman, A.R., Lohr, M., and Im, C.S. 2004. *Chlamydomonas reinhardtii* in the landscape of pigments. Annu Rev Genet 38: 119-173.

Hartshorne, J.N. 1953. The function of the eyespot in *Chlamydomonas*. New Phytol 52: 292-297.

Harz, H., Hegemann, P. 1991. Rhodopsin-regulated calcium currents in *Chlamydomonas*. Nature 351: 489-491.

Herrin, D.L., Battey, J.F., Greer, H., and Schmidt, G.W. 1992. Regulation of Chlorophyll apoprotein expression and accumulation: requirements for carotenoids and chlorophyll. J Biol Chem 267: 8260-8269.

Hirschberg, J. 2001. Carotenoid biosynthesis in flowering plants. Curr Opin Plant Biol 4: 210-218.

Hunter, W.N. Isoprenoid precursor biosynthesis offers potential targets for drug discovery against diseases caused by apicomplexan parasites. Curr Top Med Chem 11: 2048-2059.

Inwood, W., Yoshihara, C., Zalpuri, R., Kim, K.S., and Kustu. 2008. The ultrastructure of a *Chlamydomonas reinhardtii* mutant strain lacking phytoene synthase resembles that of a colorless alga. Mol Plant 1: 925-937.

Jomaa, H., Wiesner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., Türbachova et al. 1999. Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. Science 285: 1573-1576.

Krinsky, N.I. 1989. Membrane antioxidants. Ann N Y Acad Sci 551: 17-33.

Kobayashi, M., Hirai, N., Kurimura, Y., Ohigashi, H., and Tsuji, Y. 1997. Abscisic-acid dependent algal morphogenesis in the unicellular green alga *Haematococcus pluvalis*. Plant Growth Regul 22: 79-85.

Kull, D., and Pfander, H. 1995. Appendix: list of new carotenoids. In Carotenoids: Isolation and Analysis (Britton, G., Liaaen-Jensen, S., and Pfander, H., eds.) Vol 1A 295-317. Birkhäuser, Basel.

Ladygin, V.G. 2014. Biogenesis and possible modification of a carotenoid composition in the eyespot of *Chlamydomonas reinhardti*i mutants. Microbiology 83: 30-38.

Lazrak, T., Milon, A., Wolff, G., Albrecht, A.M., Mieché, M., Ourisson, G., and Nakatani, Y. 1987. Comparison of the effects of inserted C_{40} and C_{50} terminally dihydroxylated carotenoids on the mechanical properties of various phospholipid vesicles. Biochem Biophys Acta 903: 132-141.

Lell, B., Ruangweerayut, R., Wiesner, J., Missinou, M.A., Schindler, A., Baranek, T., Hintz, M., et al. 2003. Fosmidomycin, a novel chemotherapeutic agent for malaria. Antimicrob Agents Chemother 47: 735-738.

Lohr, M. 2009. Carotenoids in The Chlamydomonas Sourcebook, Organellar and Metabolic Processes. Academic Press, Canada.

Lohr, M., Im, C.S., and Grossman, A.R. 2005. Genome-based examination of chlorophyll and carotenoid biosynthesis in *Chlamydomonas reinhardti*i. Plant Physiol 138: 490-515.

Mast, S.O. 1917. The relation between spectral color and stimulation in the lower organisms. J Exp Zool 22: 471.

McCarthy, S.S., Kobayashi, M.C., and Niyogi, K.K. 2004. White mutants of *Chlamydomonas reinhardtii* are defective in phytoene synthase. Genetics 168: 1249-1257.

Nikulina, K., Chunaev, A.S., and Boschetti, A. 1999. Accumulation of z-carotene in *Chlamydomonas reinhardtii* under control of the ac5 nuclear gene. Plant Cell Rep 19: 37-42.

Niyogi, K.K. 1999. Photoprotection revisted: genetic and molecular approaches. Annu Rev Plant Physiol Plant Mol Biol 50: 333-359.

Pardha-Saradhi, S.P., Suzuki, I., Katoh, A., Sakamoto, A., Sharmila, P., Shi, D.J., and Murata, N. 2000. Protection against the photo-induced inactivation of the photosystem II complex by abscisic acid. Plant Cell Environ 23: 711-718.

Pringsheim, E.G. 1955. The Genus Polytomella. J Protozool 2: 137-145.

Ralph, S.A., van Dooren, G.G., Waller, R.F., Crawford, M.J., Fraunholz, M.J., Foth, B.J., Tonkin, D.S.R., and McFadden, G.I. 2004. Tropical infectious diseases: metabolic maps and functions of the *Plasmodium falciparum* apicoplast. Nat Rev Microbiol 2: 203-216.

Rumpf, R., Vernon, D., Schreiber, D., and Birky, C.W. Jr. 1996. Evolutionary consequences of the loss of photosynthesis in Chlamydomonadaceae: phylogenetic analysis of Rrn18 (18S rDNA) in 13 *Polytoma* strains (Chlorophyta). J Phycol 32: 119-126.

Schwender, J., Müller, C., Zeidler, J., and Lichtenthaler, H.K. 1999. Cloning and heterologous expression of a cDNA encoding 1-deoxy-D-xylulose-5-phosphate reductoisomerase of *Arabidopsis thaliana*. FEBS Lett 455:140-144.

Sineshchekov, O.A., Jung, K.H., and Spudich, J.L. 2002. Two rhodopsins mediate phototaxis to low- and high-intensity light in *Chlamydomonas reinhardtii*. Proc Natl Acad Sci USA 99: 8689-8694.

Sirevåg, R., and Levine, R.P.1973. Transcription and translation for carotenoid synthesis in *Chlamydomonas reinhardtii*. Plant 111: 73-84.

Straub, O. 1987. Key to Carotenoids. 2nd edition (Pfander, H., ed.) Birkhäuser, Basel

Sun, Z., Cunningham, F.X., and Gantt, E. 1998. Differential expression of two isopentenyl pyrophosphate isomerases and enhanced carotenoid accumulation in a unicellular chlorophyte. Proc Natl Acad Sci USA 95: 11482-11488.

Tian, L., Musetti, V., Kim, J., Magallanes-Lundback, M., and DellaPenna, D. 2004. The Arabidopsis LUT1 locus encodes a member of the cytochrome P450 family that is required for carotenoid ε-ring hydroxylation activity. PNAS 101: 402-407.

Tonhosolo, R., D'Alexandri, F.L., de Rosso, V.V., Gazarini, M.L., Matsumura, M.Y., Peres, V.J., Merino, E.F., et al. 2009. Carotenoid biosynthesis in intraerythrocytic stages of *Plasmodium falciparum*. J Biol Chem 284: 9974-9985.

Wiesner, J., Borrmann, S., and Jomaa, H. 2003. Fosmidomycin for the treatment of malaria. Parasitol Res 90: S 71-76.

Wiesner, J., Reichenberg, A., Heinrich, S., Schlitzer, M., and Jomaa, H. 2008. The plastid-like organelle of apicomplexan parasites as drug target. Curr Pharm Des 14: 855-871.

Yoshida, K., Igarashi, E., Mukai, M., Hirata, K., and Miyamoto, K. 2003. Induction of tolerance to oxidative stress in the green alga, *Chlamydomonas reinhardtii*, by abscisic acid. Plant Cell Environ 26: 451-457.

Yoshida, K., Igarashi, E., Wakatsuki, E., Miyamoto, K., and Hirata, K. 2004. Mitigation of osmotic and salt stresses by abscisic acid through reduction of stress-derived oxidative damage in *Chlamydomonas reinhardtii*. Plant Sci 167: 1335-1341.

Zhang, L.X., Conney, R.V., and Bertram, J.S. 1991. Carotenoids enhance gap junctional communication and inhibit lipid peroxidation in C3H/ 10T1/2 cells: relationship to their cancer chemopreventive action. Carcinogenesis 12: 2109-2114.

Chapter 6

6 Plastid-localized metabolic pathways and plastid complexity in *Polytomella* spp.

6.1 A summary of *Polytomella* plastid metabolic pathways

Investigation of the four *Polytomella* species transcriptome and genome data provided new insight into the diversity of plastid metabolic functions within this genus. We have shown that starch biosynthesis and degradation occurs within *Polytomella* algae, and that despite the absence of a maltose exporter, there is a diversity of glucose transporters, such as hexokinase and glucose-6-phosphate translocators. The lack of a maltose exporter may have contributed to an achlorotic plastid phenotype as well as a decreased photosynthetic rate, as have been observed to occur in mutants (Cho *et al.* 2011). Amino acid biosynthesis of over 17 amino acids was also demonstrated. Some enzymes were missing, specifically IGPS, AOD, AAH, and DPR, leading us to believe that either Illumina sequencing failed to detect their transcripts and genes, or that *Polytomella* has a new alternative and innovative way of synthesizing the amino acids that allows it to dispose of these enzymes with no major consequence to the pathway. Tetrapyrrole biosynthesis also occurs in the four *Polytomella* species; however, the pathway is retained for heme biosynthesis, since every downstream enzyme involved in chlorophyll biosynthesis is missing from both transcriptome and genome data. This finding is consistent with the importance of heme in eukaryotic organisms as well as the heterotrophic state of *Polytomella* that eliminates the need for chlorophyll and any associated function (Atteia *et al.* 2005). Maybe the most interesting of the highlighted pathways in this study is the observed carotenoid biosynthetic pathway in *Polytomella*. *P. magna*, the most basal of the *Polytomella* species, and the one most closely related to *Chlamydomonas reinhardtii*, has retained the carotenoid biosynthetic pathway; a result that is consistent with its pinkish pigmentation and retention of an eyespot structure. *P. parva* and *P. piriformis* have, on the other hand, lost the carotenoid pathway, whereas *P. capuana* is stuck in the middle with the retention of some enzymes accompanied with the loss of others leading to an incomplete carotenoid pathway. This has allowed us to conclude that carotenoid

biosynthesis loss is a result of the loss of photosynthesis, rather than the other way around as others have suggested (Inwood *et al.* 2008). Moreover, it is also our conclusion that the loss of carotenoids has occurred twice within the *Polytomella* genus; once in the branch leading to *P. capuana* and the other in the phylogenetic branch leading to *P. parva* and *P. piriformis*. *P. magna* may be the only example of a primary-plastid, nonphotosynthetic green alga that has retained a carotenoid pathway.

Table 12: Plastid metabolic pathways within nonphotosynthetic algae and *Polytomella* spp. ✓ or ✗ represents the presence or absence of the pathway, respectively.

Species	Starch	Amino Acids	Tetrapyrrole	Carotenoids
Plasmodium falciparum	✓	✗	✓	✓
Helicosporidium	✓	Limited (~6)	✓	✗
Prototheca wickerhamii	✓	Limited (~10)	✓	✗
P. magna	✓	✓	✓	✓
Polytomella spp.	✓	✓	✓	✗

6.2 *Polytomella* spp. plastid complexity

If plastid complexity was to be measured by its diversity of functions, then the four *Polytomella* species possess the most complex plastids in comparison to other nonphotosynthetic algal species, such as *Plasmodium falciparum*, *Helicosporidium* sp., and *Prototheca wickerhamii*. As shown in table 12, *P. falciparum* cannot synthesize any amino acids but instead they depend on their host's amino acid machinery. *Helicosporidium* cannot synthesize carotenoids but are able to make a total of 6 amino acids (de Koning and Keeling 2004; Pombert *et al.* 2014). Similarly, *P. wickerhamii* is unable to synthesize carotenoids and can make a total of 10 amino acids (Borza *et al.* 2005). On the other hand, *Polytomella* spp. can make more than 17 amino acids, with *P. magna* also retaining all of the enzymes required for carotenoid biosynthesis.

As can be seen in table 13 below, more enzymes are missing in *Helicosporidium* and *P. wickerhamii* in comparison to the photosynthetic *C. reinhardtii* or the free-living *Polytomella* spp. These enzymes are involved in pathways relating to tetrapyrrole, amino acid, isoprenoids, or starch biosynthesis. Some have suggested that parasitism level/ host

dependence correlates to the diversity of functions performed by the organism's plastid. Our study supports this connection as can be seen with *Polytomella* spp. increased metabolic complexity in comparison to nonphotosynthetic parasitic algal species. *Helicosporidium* and *P. wickerhamii* are both parasites with a free-living stage while *Polytomella* spp. are fully free-living. This autonomy may have contributed to a greater need for *Polytomella* algae to maintain their plastid functions.

6.3 Concluding remarks

In this study, I have addressed all of my objectives (stated in section 1.6) whereby I have shown that the plastids of *Polytomella* spp. possess all of the enzymes (nuclear-encoded, plastid-targeted) required to perform a range of functions and biosynthetic pathways. I have also failed to recover any plastid-encoded genes or transcripts (as evidenced by the absence of a plastid-encoded protochlorophyllide reductase in tetrapyrrole biosynthesis, see table 13). Moreover, I have also failed to recover any transcripts or genes related to chlorophyll biosynthesis, as shown earlier (in chapter 3 or table 13 below). These findings confirm our previous study suggesting the loss of the plastid genome within *Polytomella* algae (Smith and Asmail 2014; Smith and Lee 2014).

Nonetheless, further studies are needed to confirm my results. While I have demonstrated the presence of the genes encoding for enzymes required for these metabolic pathways; that does not guarantee that these pathways are actually occurring or that the enzymes are fully functioning. This could be done using various pathway-specific enzyme detection protocols; for instance, indole test or even functional complementation tests using *Escherichia coli* (Atteia *et al.* 2005). Some of the enzymes were absent in the transcriptome but encoded within the *Polytomella* nuclear genome. One needs to investigate whether these genes are actually being encoded but are intentionally not transcribed, or are just pseudogenes and are non functional.

Table 13: Metabolic pathways in *C. reinhardtii*, *Helicosporidium* sp., *Polytomella* spp. and *P. wickerhamii*, respectively. (For information regarding *Helicosporidium* sp. and *P. wickerhamii*, see Figueroa-Martinez *et al.* 2015 and de Koning and Keeling 2004).

	C. reinhard	*Helicosporidium*	*P.cap*	*P. mag*	*P. par*	*P. pirif*	*P. wicker*
Starch Biosynthesis							
Starch phosphorylase	✓	✓	✓	✓	✓	✓	No
Branching enzyme	✓	✓	✓	✓	✓	✓	✓
a-Glucanotransferase	✓	No	✓	✓	✓	✓	No
Starch synthase	✓	✓	✓	✓	✓	✓	No
Glucose-1-phosphate adenyltransferase	✓	✓	✓	✓	✓	✓	✓
Hexokinase	✓	No	✓	✓	✓	✓	No
Glucose-6-phosphate isomerase	✓	✓	✓	✓	✓	✓	✓
Beta-amylase	✓	No	✓	✓	✓	✓	No
Phosphoglucomutase	✓	✓	✓	✓	✓	✓	✓
Tetrapyrrole Biosynthesis							
Protochlorophyllide reductase	✓	✓	No	No	No	No	No
Coproporphyrinogen III oxidase	✓	✓	✓	✓	✓	✓	No
Protoporphyrinogen oxidase	✓	✓	✓	✓	✓	✓	No
Heme oxygenase	✓	No	✓	✓	✓	✓	No
Uroporphyrinogen III synthase	✓	✓	✓	✓	✓	✓	✓
Glutamate-1-semialdehyde-2,1-aminomutase	✓	✓	✓	✓	✓	✓	No
Glutamyl-tRNA synthetase	✓	✓	✓	✓	✓	✓	No
Glutamyl-tRNA reductase	✓	✓	✓	✓	✓	✓	No
Ferrochelatase	✓	✓	✓	✓	✓	✓	No
Uroporphyrinogen decarboxylase	✓	✓	✓	✓	✓	✓	✓
Phenylalanine, tyrosine, and tryptophan biosynthesis							
Anthranilate phosphoribosyltransferase	✓	✓	✓	✓	✓	✓	No
3-phosphoshikimate 1-carboxyvinyltransferase	✓	✓	✓	✓	✓	✓	No
Aspartate aminotransferase	✓	✓	✓	✓	✓	✓	✓
Shikimate kinase	✓	✓	✓	✓	✓	✓	No
3-deoxy-7-phosphoheptulonate synthase	✓	✓	✓	✓	✓	✓	✓

Anthranilate synthase, ANS1	✓	✓	✓	✓	✓	✓	No
Anthranilate synthase, ASB1	✓	✓	✓	✓	✓	✓	No
Tryptophan synthase, alpha, TSA	✓	✓	✓	✓	✓	✓	No
Tryptophan synthase, beta, MAA7	✓	✓	✓	✓	✓	✓	✓
3-Dehydroquinate synthase	✓	✓	✓	✓	✓	✓	✓
Chorismate synthase	✓	✓	✓	✓	✓	✓	No
Chorismate mutase	✓	✓	✓	✓	✓	✓	No
Arogenate/prephenate dehydratase	✓	✓	✓	✓	✓	✓	✓
3-Dehydroquinate dehydratase/shikimate dehydrogenase	✓	✓	✓	✓	✓	✓	No
Arogenate dehydrogenase	✓	✓	✓	✓	✓	✓	No
Glutamate/prephenate aminotransferase	✓	✓	✓	✓	✓	✓	No
Valine, Leucine, and Isoleucine Biosynthesis							
3-isopropylmalate dehydrogenase	✓	✓	✓	✓	✓	✓	No
Ketol acid reductoisomerase	✓	✓	✓	✓	✓	✓	✓
Branched-chain aminotransferase	✓	✓	✓	✓	✓	✓	No
2-isopropylmalate synthase	✓	✓	✓	✓	✓	✓	No
Acetolactate synthase, large subunit	✓	✓	✓	✓	✓	✓	✓
Acetolactate synthase, small subunit	✓	✓	✓	✓	✓	✓	✓
Dihydroxyacid dehydratase	✓	✓	✓	✓	✓	✓	No
3-isopropylmalate dehydratase, large subunit	✓	✓	✓	✓	✓	✓	No
3-isopropylmalate dehydratase, small subunit	✓	No	✓	✓	✓	✓	No
Threonine dehydratase	✓	✓	✓	✓	✓	✓	No
Isoprenoid Biosynthesis							
1-deoxy-D-xylulose-5-phosphate reductoisomerase	✓	✓	✓	✓	✓	✓	No
4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	✓	✓	✓	✓	✓	✓	No
2-C-methyl-D-erythritol 4-phosphate cytidyltransferase	✓	No	✓	✓	✓	✓	No
1-deoxy-D-xylulose-5-phosphate synthase	✓	✓	✓	✓	✓	✓	No
2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	✓	✓	✓	✓	✓	✓	No
Isopentenyl-diphosphate delta-isomerase	✓	✓	✓	✓	✓	✓	No
(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase	✓	✓	✓	✓	✓	✓	✓
Geranyl diphosphate synthase	✓	✓	✓	✓	✓	✓	✓

6.4 References

Atteia, A., van Lis, R., and Beale, S.I. 2005. Enzymes of the heme biosynthetic pathway in the nonphotosynthetic alga *Polytomella* sp. Eukaryot Cell 4: 2087-2097.

Borza, T., Popescu, C.E., and Lee, R.W. 2005. Multiple metabolic roles for the nonphotosynthetic plastid of the green alga *Prototheca wickerhamii*. Eukaryot Cell 4: 253-261.

Cho, M.H., Lim, H., Shin, D.H., Jeon, J.S., Bhoo, S.H., Park, Y.I., and Hahn, T.R. 2011. Role of the plastidic glucose translocator in the export of starch degradation products from chloroplasts in *Arabidopsis thaliana*. New Phytol 190: 101-112.

de Koning, A.P., and Keeling, P.J. 2004. Nucleus-encoded genes for plastid targeted proteins in *Helicosporidium*: functional diversity of a cryptic plastid in a parasitic alga. Eukaryot Cell 3: 1198-1205.

Figueroa-Martinez, F., Nedelcu, A.M., Smith, D.R., and Reyes-Prieto, A. 2015. When the lights go out: the evolutionary fate of free-living colorless green algae. New Phytol 206: 972-982.

Inwood, W., Yoshihara, C., Zalpuri, R., Kim, K.S., and Kustu. 2008. The ultrastructure of a *Chlamydomonas reinhardtii* mutant strain lacking phytoene synthase resembles that of a colorless alga. Mol Plant 1: 925-937.

Pombert, J.F., Blouin, N.A., Lane, C., Boucias, D., and Keeling, P.J. 2014. A lack of parasitic reduction in the obligate parasitic green alga *Helicosporidium*. PLOS Genet 10:e1004335. doi:10.1371/journal.pgen.1004355.

Smith, D.R., and Asmail, S.R. 2014. Next-generation sequencing data suggest that certain nonphotosynthetic green plants have lost their plastid genomes. New Phytol 204: 7-11.

Smith, D.R., and Lee, R.W. 2014. A plastid without a genome: evidence from the nonphotosynthetic green algal genus *Polytomella*. Plant Physiol 164: 1812-1819.

Appendices

Appendix A: Next-generation sequencing data suggest that certain nonphotosynthetic green plants have lost their plastid genomes.

New Phytologist

Forum

Letters

Next-generation sequencing data suggest that certain nonphotosynthetic green plants have lost their plastid genomes

Introduction

Genomes are the agents of life; they are present, in one form or another, in all living things, and the latter cannot exist without the former. Sometimes, however, genomes exist long after relinquishing control of the 'life' that they once yielded. Take, for instance, the mitochondrion and plastid of eukaryotic cells. Both descend from once free-living bacteria, which, over millions of years, slowly surrendered their autonomy to the host cell that engulfed them. Nonetheless, mitochondria and plastids still contain a genome and gene expression system, even though almost all of the proteins required for these two organelles to function are nuclear encoded – a consequence of massive, recurring waves of organelle-to-nucleus gene migration (Kleine *et al.*, 2009). Why, then, have not mitochondria and plastids outsourced all of their genes to the nucleus? Why do they retain a genome and a gene expression infrastructure? Satisfactory answers to these questions, particularly for anaerobic or nonphotosynthetic species, are for the most part lacking, but accumulating data from diverse lineages suggest that some organelles have, in fact, jettisoned their genomes.

The most convincing evidence for organelle genome loss comes from the mitochondrial-derived organelles of various microbes living in anoxic environments (Hjort *et al.*, 2010). In disparate groups across the eukaryotic tree, mitochondria have been subverted into anaerobic organelles – called hydrogenosomes or mitosomes – which no longer perform oxidative phosphorylation, but continue to carry out crucial cellular processes, such as hydrogen production (de Graaf & Hackstein, 2012). Although the vestiges of a mitochondrial chromosome (mtDNA) exist in some hydrogenosomes (Pérez-Brocal & Clark, 2008; de Graaf *et al.*, 2011), many, and perhaps most, anaerobic mitochondrial-derived organelles have disposed of their DNA entirely and rely solely on nuclear-encoded proteins to function (Hjort *et al.*, 2010). The search for plastids without DNA, however, has been less fruitful.

The preservation of plastid DNA

Plastids are found in almost every ecosystem on the planet. Since their archaeplastidal origin through the primary endosymbiosis of a cyanobacterium, plastids have subsequently spread, through eukaryotic–eukaryotic endosymbioses, to remote eukaryotic groups (Archibald, 2009). Consequently, a significant proportion of the known eukaryotic diversity contains a plastid, and with a few potential exceptions (see later), wherever a plastid exists, a genome persists (Keeling, 2010). Even plastids that have lost photosynthetic capabilities, such as those of malaria parasites and heterotrophic plants and algae, are consistently shown to have a genome, albeit one that is highly reduced (*c.* 30–100 kb) with a much smaller gene content than their counterparts in closely related photosynthetic taxa (Wilson *et al.*, 1996; de Koning & Keeling, 2006; Wicke *et al.*, 2013). Given the ubiquity of plastid DNA (ptDNA) across plastid-bearing taxa, it has been argued that plastids are irreversibly tied to their genomes (Barbrook *et al.*, 2006a; Nair & Striepen, 2011). Nevertheless, certain eukaryotic lineages are believed to have lost their plastids outright (Keeling, 2010), implying that on the way to a plastid-less state there was a transitional stage in which there existed plastid-containing species without ptDNA. Some believe that it is just a matter of time until someone stumbles upon such species (Nickrent *et al.*, 1997; Palmer, 1997).

Well, that time may have come. Two recent studies, published within one month of each other, provided evidence for plastid genome loss in distinct nonphotosynthetic green plants. One is of the parasitic flowering plant *Rafflesia* (Molina *et al.*, 2014) and the other is of the colorless green alga *Polytomella* (Smith & Lee, 2014). Both investigations used next-generation sequencing (NGS) results to argue for the absence of ptDNA.

Next-generation organelle-genome sequencing

High-throughput sequencing methods have transformed the field of organelle genomics making it fast, easy, and efficient to sequence mtDNA and ptDNA (Smith, 2012). A single run of whole genomic DNA from a plant or alga on an NGS platform typically yields enough organelle-derived sequences to assemble complete mitochondrial and plastid genomes with > 100-fold coverage (Nock *et al.*, 2011; McPherson *et al.*, 2013). In many cases, 5–25% of the reads generated from high-throughput sequencing of total eukaryotic DNA (or RNA) come from organelles (Smith, 2012), with plastid-derived reads often outnumbering mitochondrial ones (Molina *et al.*, 2014). This is also true for nonphotosynthetic plants and algae, many of which have had their ptDNAs sequenced using next-generation techniques (Arisue *et al.*, 2012; Wicke *et al.*, 2013; Imura *et al.*, 2014). Time and again, NGS of nonphotosynthetic, plastid-bearing species has returned prodigious amounts of ptDNA data, so when researchers carried out intensive Illumina sequencing of *Rafflesia* and *Polytomella*, one would have expected them to uncover an abundance of plastid-derived reads. But they found the opposite.

New Phytologist

Rafflesia: big flower, but no ptDNA?

Rafflesia is a southeast Asian genus of angiosperms, situated within the rosids, and sometimes called 'corpse flower'. Credited for having the largest single flower of any plant, it is comprised of putrid-smelling, nonphotosynthetic parasites, which lack stems, roots, and leaves, and rely solely on their host, the vine *Tetrastigma*, for survival (Nais, 2001). *Rafflesia* is just one of many land plant genera that harbor nonphotosynthetic, parasitic species; others include *Cuscuta*, *Epifagus*, and *Orobanche*, to name but a few – see Westwood *et al.* (2010) for a review on the topic.

Over the past 25 yr, biologists have identified and sequenced ptDNA from a wide range of holoparasitic plants (Krause, 2011), such as *Epifagus virginiana* and *Orobanche gracilis* (Wolfe *et al.*, 1992; Wicke *et al.*, 2013), strengthening the idea that nonphotosynthetic plastids require a genome. However, PCR and Southern blot experiments failed to identify ptDNA in members of the *Rafflesia* genus, hinting that this nonphotosynthetic lineage may lack a plastid genome (Nickrent *et al.*, 1997; Davis *et al.*, 2007). *Rafflesia* mtDNA, on the other, has been highly amenable to sequencing and is well characterized (Xi *et al.*, 2013).

Scientists are now stepping up the search for *Rafflesia* ptDNA. Molina *et al.* (2014) isolated whole genomic DNA from a *Rafflesia lagascae* floral bud, collected in Cagayan province, Philippines, and subjected it to Illumina sequencing. The resulting 440 million paired-end reads were teeming with mtDNA but contained very few ptDNA-like sequences, none of which appeared to come from the *R. lagascae* plastid. Of the *c.* 1.5 million *R. lagascae* Illumina contigs, only *c.* 45 (0.003%; 11.5 kb) showed similarity to genic and intergenic sequences typically found in land plant plastid genomes. But not one of these plastid contigs contained a complete gene or an intact open reading frame, and they all had low read coverage (*c.* 1.5×), contrasting the >300-fold coverage observed for the mtDNA-derived contigs. Moreover, none of the plastid sequences were found to be phylogenetically associated with close relatives of *Rafflesia*, such as *Ricinus* or *Hevea*, and many were affiliated with species closely related to *Tetrastigma* – the plant that *R. lagascae* parasitizes (Fig. 1). Based on these findings, the authors argue that plastid sequences recovered from *R. lagascae* Illumina sequencing are nuclear-located (and in a few cases mitochondrial-located) ptDNA-like sequences, which have been horizontally transferred to *R. lagascae* from the plastid genome of *Tetrastigma*. Host-to-parasite horizontal gene transfer is well documented in angiosperms (Davis & Wurdack, 2004), and more than a quarter of the mtDNA-encoded genes in *Rafflesia* species, including *R. lagascae*, appear to originate from *Tetrastigma* (Xi *et al.*, 2013; Molina *et al.*, 2014). If *R. lagascae* does have a plastid genome it likely is in a cryptic form, has a highly divergent sequence, and/or is found at very low levels. Or perhaps, as the authors suggested, it has vanished altogether. If so, it may not be alone. Similar experiments have indicated that another lineage from the Viridiplantae may have also discarded its ptDNA.

Potential plastid genome loss in *Polytomella* green algae

First described over a century ago, *Polytomella* is a monophyletic green algal genus of free-living, freshwater unicells, closely related to the model photosynthetic species *Chlamydomonas reinhardtii* and *Volvox carteri* (Pringsheim, 1955; Smith *et al.*, 2013) (Fig. 1). Although nonphotosynthetic, *Polytomella* algae have a plastid (Moore *et al.*, 1970), but early attempts to identify a gene expression system within it were unsuccessful (Nedelcu *et al.*, 1996; Nedelcu, 2001), even though similar techniques identified one in the plastids of other colorless green algae, such as *Prototheca*

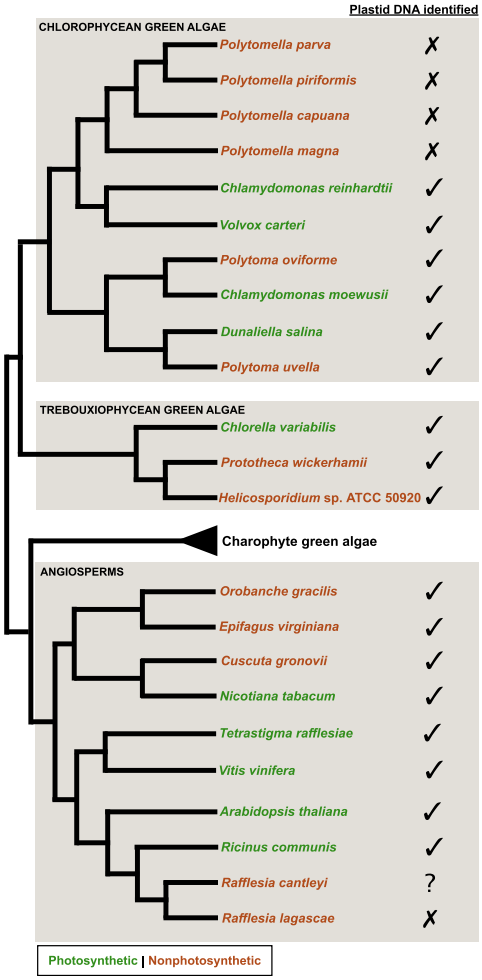


Fig. 1 Tree of chlorophycean and trebouxiophycean green algae and angiosperms showing examples of species that have lost photosynthetic capabilities. Photosynthetic species, green text; nonphotosynthetic, red text. Branching order based on published phylogenetic analyses (Nedelcu, 2001; Smith *et al.*, 2013; Xi *et al.*, 2013, and references cited therein).

wickerhamii and *Polytoma uvella*, both of which lost photosynthesis independently of *Polytomella* (Fig. 1).

Recently, Smith & Lee (2014) used high-throughput sequencing to search for ptDNA and plastid gene expression in *Polytomella*. Illumina sequencing and assemblies of total DNA isolated from each of the four known *Polytomella* lineages (*P. parva*, *P. piriformis*, *P. capuana*, and *P. magna*) (Fig. 1) gave > 225 million paired-end reads and 200 Mb of contig sequences. These data were scanned using BLAST- and mapping-based methods for putative *Polytomella* ptDNA sequences, but none were found. The same methods, however, easily detected *Polytomella* mtDNA-derived reads and contigs, despite the fact that *Polytomella* mitochondrial genomes have highly reduced gene contents, elevated rates of nucleotide substitution, fragmented architectures, and/or extreme nucleotide compositions (Smith *et al.*, 2010a).

Illumina RNA sequencing (RNA-seq) and transcriptomic analysis of *P. parva* also provided no signs of a plastid genome or associated gene expression system (Smith & Lee, 2014). Assembly of *c.* 50 million RNA-seq reads and annotation of *c.* 31 000 contigs uncovered thousands of *P. parva* nuclear transcripts, hundreds of which code for putative plastid-targeted proteins. Close inspection of these presumed plastid proteins indicated that the *P. parva* plastid performs a diversity of functions, similar to those observed in the plastids of other nonphotosynthetic algae (Borza *et al.*, 2005). Conspicuously absent, however, were any potential plastid-targeted proteins involved in the expression, replication, or repair of ptDNA, such as plastid-like ribosomal proteins. These data, along with the inability to detect ptDNA-derived sequencing reads, ultimately led Smith & Lee (2014) to conclude that the *Polytomella* plastid genome is nonexistent.

Proof of ptDNA absence or absence of ptDNA proof?

It is possible that a cryptic plastid genome is hiding within *R. lagascae* and *Polytomella* algae and that it somehow escaped detection by high-throughput sequencing. Illumina sequencing has its drawbacks: it has been shown to give uneven and poor read coverage across genomic regions with extremely biased base compositions (Oyola *et al.*, 2012), which could impede the identification of a possibly small plastid genome. Moreover, plastid genomes can sometimes have peculiar architectures, such as fragmented chromosomes (Barbrook *et al.*, 2006b), large numbers of introns and repetitive DNA (Smith *et al.*, 2010b), and/or high levels of post-transcriptional editing (Tillich *et al.*, 2006) – features that could hinder the identification of ptDNA via NGS methods.

But even when considering the potential issues associated with NGS and plastid genome architecture, nuclear transcriptome sequencing should still provide evidence of ptDNA expression, replication, and repair. In the case of *P. parva*, transcriptomic analyses revealed no nuclear-encoded, plastid-targeted proteins with ptDNA-related functions, which is consistent with plastid genome loss. For *R. lagascae*, unfortunately, there are currently no published data on nuclear transcripts for plastid-targeted proteins. But within the National Center for Biotechnological Information Sequence Read Archive there are 4.4 Gb of paired-end Illumina RNA-seq data for *R. cantleyi* (accession number SRX157681), which is a close relative of *R. lagascae*. Searching these RNA-seq reads for nuclear transcripts encoding plastid proteins should be straightforward, and is a critical step in the pursuit of a *Rafflesia* plastid genome. If *Rafflesia* has ptDNA there should be dozens of nuclear-encoded proteins with ptDNA-related functions.

Presently, the evidence for plastid genome loss in *R. lagascae* and *Polytomella* are based solely on the results of NGS experiments (Molina *et al.*, 2014; Smith & Lee, 2014) as well as some preliminary PCR and/or nucleotide hybridization work (Nedelcu *et al.*, 1996; Nickrent *et al.*, 1997; Nedelcu, 2001; Davis *et al.*, 2007). Further explorations for a potential plastid genome in these lineages could come from fluorescent microscopy using DNA-binding dyes, such as DAPI or SYBR Green as well as from additional analyses of the available *Rafflesia* and *Polytomella* NGS data – specifically, analyses using different assembly, BLAST, and mapping approaches than those employed by Molina *et al.* (2014) and Smith & Lee (2014). If ptDNA does exist in *Rafflesia* and/or *Polytomella* species, fluorescent microscopy should reveal nucleoids within their plastids. However, an attempt to do such an experiment in *P. parva* was complicated by highly reticulated mitochondrial structures, which layered over and obscured potential nucleoid signals from the plastid (Smith & Lee, 2014). Arguably the strongest evidence for plastid genome loss will come from complete nuclear genome sequencing of various *Rafflesia* and *Polytomella* taxa, which should provide a complete suite of nuclear genes for plastid-targeted proteins and consequently a better understanding of plastid function within these species.

If plastid genome loss has occurred in the *Rafflesia* and *Polytomella* lineages, one will need to explain how they are synthesizing heme. In most plastid-bearing species heme biosynthesis begins in the plastid, via the C_5 pathway, and employs a plastid-encoded tRNA glutamate (Beale, 1999). It has been hypothesized that nonphotosynthetic plants and algae retain a plastid genome, and its $tRNA^{Glu}$, to maintain a functional heme pathway (Barbrook *et al.*, 2006a). In some plastid-bearing species, including the malaria parasite *Plasmodium falciparum,* the initial steps of heme biosynthesis occur in the mitochondrion through the Shemin pathway (Oborník & Green, 2005), and they are therefore not reliant on a plastid $tRNA^{Glu}$. It is not known how *Rafflesia* or *Polytomella* are synthesizing heme, but the latter, at least, does not appear to be using the Shemin pathway (Smith & Lee, 2014).

More proposed cases of plastid genome loss on the way?

There are a large number of poorly studied eukaryotic microbial groups, many of which harbor species with nonphotosynthetic plastids or potentially with unidentified 'cryptic' plastids (Keeling, 2010), including various lineages within the eukaryotic superphylum Alveolata, such as colpodellids (Gile & Slamovits, 2014), perkinsids (Robledo *et al.*, 2011), and *Oxyrrhis* (Slamovits & Keeling, 2008). As researchers explore these groups they will likely discover more possible examples of plastid genome loss. It is surprising that currently the two best cases for ptDNA loss – *Rafflesia* and *Polytomella* – come from lineages whose plastids descend directly from a primary endosymbiosis of a

cyanobacterium and not from those whose plastids derive from eukaryote–eukaryote endosymbioses. It is the latter category of plastids that are thought to have been lost completely (genome and all) in certain protist groups, such as *Cryptosporidium* (Keeling, 2010), whereas there are no purported examples of outright plastid loss in any primary plastid-bearing lineages. There is mounting evidence that the oyster parasite *Perkinsus marinus* (Alveolata) has a relic, red-algal-derived plastid without a genome (Robledo *et al.*, 2011). And there may well be other protists with undiscovered relic plastids without genomes. However, the next case for plastid genome loss could come from another land plant. A survey of ptDNA gene content across the parasitic plant genus *Cuscuta* identified, through nucleotide-hybridization work, some species that may have lost their plastid genomes (Braukmann *et al.*, 2013).

It is perplexing to imagine the steps involved in acquiring a photosynthetic organelle – from the endosymbiosis of a free-living photosynthetic organism to the integration of that symbiont into the host cell to the amalgamation of symbiont and host genomes. It is equally perplexing to envision the reverse process – the forfeiting of photosynthetic capabilities, the deterioration of a plastid genome, and the eventual loss of the organelle itself. Either way, both of these processes have a lot to teach us about the evolution and diversity of life.

Acknowledgements

The authors wish to thank Susann Wicke and two anonymous reviewers for their critical reading of the manuscript and helpful feedback. D.R.S. is supported by a Discovery Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

David Roy Smith* and Sara Raad Asmail

Department of Biology, University of Western Ontario, London, ON, N6A 5B7, Canada
(*Author for correspondence: tel +1 519 661 2111, ext 86482; email dsmit242@uwo.ca)

References

Archibald JM. 2009. The puzzle of plastid evolution. *Current Biology* **19**: R81–R88.

Arisue N, Hashimoto T, Mitsui H, Palacpac NM, Kaneko A, Kawai S, Hasegawa M, Tanabe K, Horii T. 2012. The *Plasmodium* apicoplast genome: conserved structure and close relationship of *P. ovale* to rodent malaria parasites. *Molecular Biology and Evolution* **29**: 2095–2099.

Barbrook AC, Howe CJ, Purton S. 2006a. Why are plastid genomes retained in non-photosynthetic organisms? *Trends in Plant Science* **11**: 101–108.

Barbrook AC, Santucci N, Plenderleith LJ, Hiller RG, Howe CJ. 2006b. Comparative analysis of dinoflagellate chloroplast genomes reveals rRNA and tRNA genes. *BMC Genomics* **7**: 297.

Beale SI. 1999. Enzymes of chlorophyll biosynthesis. *Photosynthesis Research* **60**: 43–73.

Borza T, Popescu CE, Lee RW. 2005. Multiple metabolic roles for the nonphotosynthetic plastid of the green alga *Prototheca wickerhamii*. *Eukaryotic Cell* **4**: 253–261.

Braukmann T, Kuzmina M, Stefanovic S. 2013. Plastid genome evolution across the genus *Cuscuta* (Convolvulaceae): two clades within subgenus *Grammica* exhibit extensive gene loss. *Journal of Experimental Botany* **64**: 977–989.

Davis CC, Latvis M, Nickrent DL, Wurdack KJ, Baum DA. 2007. Floral gigantism in Rafflesiaceae. *Science* **315**: 1812.

Davis CC, Wurdack KJ. 2004. Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. *Science* **305**: 676–678.

Gile GH, Slamovits CH. 2014. Transcriptomic analysis reveals evidence for a cryptic plastid in the colpodellid *Voromonas pontica*, a close relative of chromerids and apicomplexan parasites. *PLoS ONE* **9**: e96258.

de Graaf RM, Hackstein JHP. 2012. Hydrogenosomes and mitosomes: mitochondrial adaptations to life in anaerobic environments. In: Altenbach AV, Bernard JM, Seckbach J, eds. *Anoxia: evidence for eukaryote survival and paleontological strategies, cellular origin, life in extreme habitats and astrobiology, vol. 21*. New York, NY, USA: Springer, 83–112.

de Graaf RM, Ricard G, van Alen TA, Duarte I, Dutilh BE, Burgtorf C, Kuiper JW, van der Staay GW, Tielens AGM, Huynen MA *et al.* 2011. The organellar genome and metabolic potential of the hydrogen-producing mitochondrion of *Nyctotherus ovalis*. *Molecular Biology and Evolution* **28**: 2379–2391.

Hjort K, Goldberg AV, Tsaousis AD, Hirt RP, Embley TM. 2010. Diversity and reductive evolution of mitochondria among microbial eukaryotes. *Philosophical Transactions of the Royal Society B* **365**: 713–727.

Imura T, Sato S, Sato Y, Sakamoto D, Isobe T, Murata K, Holder AA, Yukawa M. 2014. The apicoplast genome of *Leucocytozoon caulleryi*, a pathogenic apicomplexan parasite of the chicken. *Parasitology Research* **113**: 823–828.

Keeling PJ. 2010. The endosymbiotic origin, diversification and fate of plastids. *Philosophical Transactions of the Royal Society B* **365**: 729–748.

Kleine T, Maier UG, Leister D. 2009. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annual Review of Plant Biology* **60**: 115–138.

de Koning AP, Keeling PJ. 2006. The complete plastid genome sequence of the parasitic green alga *Helicosporidium* sp. is highly reduced and structured. *BMC Biology* **4**: 12.

Krause K. 2011. Piecing together the puzzle of parasitic plant plastome evolution. *Planta* **234**: 647–656.

McPherson H, van der Merwe M, Delaney SK, Edwards MA, Henry RJ, McIntosh E, Rymer PD, Milner ML, Siow J, Rossetto M. 2013. Capturing chloroplast variation for molecular ecology studies: a simple next generation sequencing approach applied to a rainforest tree. *BMC Ecology* **13**: 8.

Molina J, Hazzouri KM, Nickrent D, Geisler M, Meyer RS, Pentony MM, Flowers JM, Pelser P, Barcelona J, Inovejas SA *et al.* 2014. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). *Molecular Biology and Evolution* **31**: 793–803.

Moore J, Cantor MH, Sheeler P, Kahn W. 1970. The ultrastructure of *Polytomella agilis*. *Journal of Protozoology* **17**: 671–676.

Nair SC, Striepen B. 2011. What do human parasites do with a chloroplast anyway? *PLoS Biology* **9**: e1001137.

Nais J. 2001. *Rafflesia of the world.* Kota Kinabalu, Borneo: Natural History Publications.

Nedelcu AM. 2001. Complex patterns of plastid 16S rRNA gene evolution in nonphotosynthetic green algae. *Journal of Molecular Evolution* **53**: 670–679.

Nedelcu AM, Spencer DF, Denovan-Wright EM, Lee RW. 1996. Discontinuous mitochondrial and chloroplast large subunit ribosomal RNAs among green algae: phylogenetic implications. *Journal of Phycology* **32**: 103–111.

Nickrent DL, Ouyang Y, Duff RJ, dePamphilis CW. 1997. Do nonasterid holoparasitic flowering plants have plastid genomes? *Plant Molecular Biology* **34**: 717–729.

Nock CJ, Waters DLE, Edwards MA, Bowen SG, Rice N, Cordeiro GM, Henry RJ. 2011. Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnology Journal* **9**: 328–333.

Oborník M, Green BR. 2005. Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. *Molecular Biology and Evolution* **22**: 2343–2353.

Oyola SO, Otto TD, Gu Y, Maslen G, Manske M, Campino S, Turner DJ, MacInnis B, Kwiatkowski DP, Swerdlow HP *et al.* 2012. Optimizing illumina next-generation sequencing library preparation for extremely at-biased genomes. *BMC Genomics* **13**: 1.

Palmer JD. 1997. Organelle genomes: going, going, gone!. *Science* **275**: 790–791.

Pérez-Brocal V, Clark CG. 2008. Analysis of two genomes from the mitochondrion-like organelle of the intestinal parasite *Blastocystis*: complete sequences, gene content, and genome organization. *Molecular Biology and Evolution* **25**: 2475–2482.

Pringsheim EG. 1955. The genus *Polytomella*. *Journal of Protozoology* **2**: 137–145.
Robledo JAF, Caler E, Matsuzaki M, Keeling PJ, Shanmugam D, Roos DS, Vasta GR. 2011. The search for the missing link: a relic plastid in *Perkinsus*? *International Journal of Parasitology* **41**: 1217–1229.
Slamovits CH, Keeling PJ. 2008. Plastid-derived genes in the nonphotosynthetic alveolate *Oxyrrhis marina*. *Molecular Biology and Evolution* **25**: 1297–1306.
Smith DR. 2012. Not seeing the genomes for the DNA. *Briefings in Functional Genomics* **11**: 289–290.
Smith DR, Hua J, Archibald JM, Lee RW. 2013. Palindromic genes in the linear mitochondrial genome of the nonphotosynthetic green alga *Polytomella magna*. *Genome Biology and Evolution* **5**: 1661–1667.
Smith DR, Hua J, Lee RW. 2010a. Evolution of linear mitochondrial DNA in three known lineages of *Polytomella*. *Current Genetics* **56**: 427–438.
Smith DR, Lee RW. 2014. A plastid without a genome: evidence from the nonphotosynthetic green alga *Polytomella*. *Plant Physiology* **164**: 1812–1819.
Smith DR, Lee RW, Cushman JC, Magnuson JK, Tran D, Polle JE. 2010b. The *Dunaliella salina* organelle genomes: large sequences, inflated with intronic and intergenic DNA. *BMC Plant Biology* **10**: 83.
Tillich M, Lehwark P, Morton BR, Maier UG. 2006. The evolution of chloroplast RNA editing. *Molecular Biology and Evolution* **23**: 1912–1921.
Westwood JH, Yoder JI, Timko MP, dePamphilis CW. 2010. The evolution of parasitism in plants. *Trends in Plant Science* **15**: 227–235.
Wicke S, Muller KF, dePamphilis CW, Quandt D, Wickett NJ, Zhang Y, Renner SS, Schneeweiss GM. 2013. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell* **25**: 3711–3725.
Wilson RJM, Denny PW, Preiser PR, Rangachari K, Roberts K, Roy A, Whyte A, Strath M, Moore DJ, Moore PW *et al.* 1996. Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *Journal of Molecular Biology* **261**: 155–172.
Wolfe KH, Morden CW, Palmer JD. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proceedings of the National Academy of Sciences, USA* **89**: 10648–10652.
Xi Z, Wang Y, Bradley RK, Sugumaran M, Marx CJ, Rest JS, Davis CC. 2013. Massive mitochondrial gene transfer in a parasitic flowering plant clade. *PLoS Genetics* **9**: e1003265.

Key words: chloroplast genome, next-generation sequencing (NGS), nonphotosynthetic green plants, organelle DNA, photosynthesis, *Polytomella*, *Rafflesia*.

Appendix B: Retention, erosion, and loss of the carotenoid biosynthetic pathway in the nonphotosynthetic green algal genus *Polytomella*. Paper submission to *New Phytologist*, along with all of its supplementary data, tables, and figures.

Retention, erosion, and loss of the carotenoid biosynthetic pathway in the nonphotosynthetic green algal genus *Polytomella*.

Letter

Sara Raad Asmail[1], David Roy Smith[1,*]

[1]Department of Biology, University of Western Ontario, London, ON, Canada, N6A 5B7

[*]Corresponding author:
David Roy Smith
Department of Biology
University of Western Ontario
London, Ontario, N6A 5B7, Canada.
Email: dsmit242@uwo.ca
Tel: (519) 661 2111, ext.86482

Key words: beta-carotene; carotenoid; *Chlamydomonas*; chloroplast; eyespot; photosynthesis; phytoene synthase.

Running header: Carotenoid biosynthesis in *Polytomella*.

Total word count: 1,790.
Number of references: 21.
Number of Tables/Figures: 1 Figure (colour)
Supporting Information: Yes.

Introduction

The evolutionary loss of carotenoid biosynthesis is often tied to the loss of photosynthesis, which is not surprising. In plants and algae, carotenoids are primarily associated with photosynthetic processes, from light absorption to photosystem assembly to protection from photodamage (Lohr, 2009; Cazzonelli, 2011; Santabarbara *et al*., 2013), and they are also key constituents of algal eyespots—specialized, typically plastid-located optical devices that detect light and direct phototaxis (Kreimer, 2009; Ladygin & Semenova, 2014). In fact, a nonphotosynthetic mutant of the model green alga *Chlamydomonas reinhardtii* lacking phytoene synthase—one of the first enzymes in carotenoid biosynthesis (Fig. 1a)—bears a remarkable resemblance to naturally occurring colourless algae, exhibiting starch accumulation, a disorganized eyespot, and no pyrenoid (Inwood *et al*., 2008). This observation has lead some to suggest that mutations to the carotenoid pathway could be responsible for the evolution of nonphotosynthetic algae (Inwood *et al*., 2008), many of which are missing the genes for carotenoid production (Borza *et al.,* 2005; Pombert *et al.,* 2014; Figueroa-Martinez *et al*., 2015), with some notable exceptions (Tonhosolo *et al*., 2009).

One algal system that could prove particularly useful for investigating the relationship between carotenoid biosynthesis and a heterotrophic existence is *Polytomella*: a monophyletic genus of nonphotosynthetic, free-living unicells closely related to *C. reinhardtii* (Pringsheim, 1955; Smith *et al*., 2013), and not to be confused with the polyphyletic genus *Polytoma*, which lost photosynthesis independently of *Polytomella* (Figueroa-Martinez *et al*., 2015). There are currently four well-described *Polytomella* lineages (Fig. 1b), represented by *P. magna*, *P. capuana*, *P. piriformis*, and *P. parva* (Göttingen Culture Collection of Algae, SAG, strains 63-9, 63-5, 63-10, and 63-3, respectively) (Smith *et al.*, 2013). The three latter species have no discernable eyespot and are white in colour, whereas *P. magna*, the deepest-branching of the four species, has an eyespot and is pinkish (Pringsheim, 1955; Smith *et al.*, 2013; MacDonald & Lee, 2014).

Recently, our lab group along with collaborators generated large amounts of Illumina RNA and DNA sequencing data from all four *Polytomella* lineages, which can be found in GenBank's Sequence Read Archive (accession numbers SRX363995,

SRX377397, SRX377560, SRX551283, and SRX710730–2). These data have already been used to explore certain plastid-located pathways and to argue that *Polytomella* algae do not have a plastid genome (Smith & Lee, 2014). In our ongoing work towards developing a complete and polished *Polytomella* genome and transcriptome, we discovered an interesting feature regarding the carotenoid biosynthetic pathway (described below), which to the best of our knowledge has heretofore not been observed in any other group of nonphotosynthetic or photosynthetic algae.

Presence and progressive loss of the carotenoid pathway in *Polytomella* algae

The genetic architecture of the carotenoid pathway in *C. reinhardtii* is well characterized (Lohr *et al*. 2005; Lohr 2009) and involves a series of enzymatic steps, all of which occur within the plastid, downstream of the non-mevalonate methylerythritol phosphate (MEP) pathway of isoprenoid biosynthesis (Fig. 1a). In an effort to identify these same pathways in *Polytomella*, we blasted the nucleotide and deduced amino acid sequences of the *C. reinhardtii* MEP and carotenoid genes against the draft transcriptome and genome assemblies of *P. magna*, *P. capuana*, *P. piriformis*, and *P. parva* (Fig. 1C; Supporting Information Methods S1 and Table S1). Full-length transcripts and genes for all of the MEP enzymes were easily identified in each of the four *Polytomella* species (Fig. 1c; Supporting Information Table S1), indicating that the MEP pathway is intact and functional. The presence and absence of genes for carotenoid enzymes, however, varied across the genus.

For *P. magna*, we recovered complete transcript and gene sequences for all of the enzymes in the carotenoid pathway up to and including lycopene β-cyclase (LCYB), which synthesizes β-carotene through the cyclization of lycopene (Fig. 1a & 1c). This finding is consistent with *P. magna* having an eyespot—an organelle that is known to be rich in β-carotene (Ladygin & Semenova, 2014). Entire *P. magna* transcripts and genes were also collected for carotene β-ketolase (BKT) and carotene β-hydroxylase (CHYB), which are involved in canthaxanthin and zeaxanthin synthesis, respectively. The nucleotide sequences for the remaining carotenoid enzymes, however, were not found in either the transcriptome or genome of *P. magna*, implying that its carotenoid pathway ends shortly downstream of LCYB (Fig. 1c).

In contrast to *P. magna*, the entire carotenoid pathway appears to be missing from both *P. parva* and *P. piriformis*, which are the two most derived and closely related of the four *Polytomella* species explored here (Fig. 1b). Neither transcripts nor genes for any of the carotenoid enzymes could be identified from *P. parva* or *P. piriformis* (Fig. 1c), which is inline with their lack of both pigmentation and eyespot structures (Pringsheim, 1955; MacDonald & Lee, 2014). Pseudogenes showing resemblance to carotenoid enzymes were also undetectable in these two taxa.

The most unexpected observations came from *P. capuana*: the genetic architecture of its carotenoid pathway is intermediate to those of *P. magna* and *P. parva*/*P. piriformis*. Like with *P. magna*, we uncovered *P. capuana* transcript and gene sequences for each of the carotenoid enzymes leading to and including LCYB and CHYB (Fig. 1c; Supporting Information Table S1). But unlike *P. magna*, the nucleotide and putative amino acid sequences of the *P. capuana* transcripts and genes were abnormal, containing large insertions and having very poor sequence identity relative to the carotenoid transcripts/genes of *P. magna*, *C. reinhardtii*, and *Volvox carteri*, and in three cases they had premature stop codons (Fig. 1c & 1d). Moreover, the enzyme BKT, which was present in the transcriptome of *P. magna*, was missing from that of *P. capuana*, but a likely BKT pseudogene was discovered in the *P. capuana* genome (Supporting Information Table S1). Together, these data suggest that the *P. capuana* carotenoid pathway is in a state of decay.

***P. capuana* genes for carotenoid biosynthesis: another one bites the dust**

Close inspection of the *P. capuana* molecular sequence data indicates that the genes encoding carotenoid enzymes are under relaxed selective constraints and are accumulating deleterious mutations (Fig. 1d; Supporting Information Table S1). Indeed, three of the seven recovered transcripts contained premature stop codons (Fig. 3c) and six contained one or more large (>50 amino acid) insertions; these same features were also observed in the corresponding *P. capuana* genomic sequences. One of the most extreme examples of genetic degeneration within the *P. capuana* carotenoid pathway is the transcript representing LCYB: not only does it have an internal stop codon, but its putative coding sequence—again, because of a series of large insertions—is potentially

>1000 nt longer than the LCYB transcripts from *P. magna*, *C. reinhardtii*, and *V. carteri* (Fig. 1d; Supporting Information Table S1). Likewise, the *P. capuana* CHYB transcript contained so many insertions that it was not possible to accurately align it to those from other chlamydomonadalean algae (Supporting Information Table 1S). Signs of relaxed selection were also observed in the CHYB and BKT genes of *P. magna*, both of which catalyze reactions at the very end of its carotenoid pathway (Fig. 1c & 1d). Similar to the carotenoid genes from *P. capuana*, the *P. magna* BKT coding sequence had three large (>150 nt) insertions and was more than 600 nt longer than the BKT transcripts of *C. reinhardtii* and *V. carteri* (Fig. 1d; Supporting Information Table S1).

The aberrant carotenoid gene sequences in *P. capuana* and *P. magna*, and the complete loss of these genes from *P. parva* and *P. piriformis* signify that the carotenoid pathway in *Polytomella* is at various stages of degradation and loss, which is presumably linked to the presence or absence of an eyespot among its members. What's more, this degradation follows a phylogenetic pattern whereby the pathway is present in the deepest branching lineage (*P. magna*), lost in the most derived one (*P. parva/P. piriformis*), and is in an intermediate stage of loss in *P. capuana*, which branches between the former two lineages (Fig. 1b). This pattern could make *Polytomella* an attractive group for studying the evolution, function, and loss of carotenoid biosynthesis in algae.

***Polytomella*: a model system for studying the retention and loss of carotenoids**

Research into carotenoids has major implication for medicine, industry, and evolution (Cazzonelli, 2011; Shumskaya & Wurtzel, 2013). Carotenoids are crucial for human health, providing precursors for vitamin A biosynthesis, but they need to be acquired through diet, which has led to the genetic engineering of β-carotene-rich crops, such as "golden rice" (Ye *et al*., 2000). Carotenoids are also manufactured on an industrial scale for use in nutritional supplements, medicines, and cosmetics (Cazzonelli, 2011). Among the best-studied organisms for synthesizing, harvesting, and genetically modifying carotenoids are chlamydomonadalean algae (e.g., Cordero *et al*., 2011), including *Haematococcus pluvialis* and *Dunaliella salina* (Fassett & Coombes, 2011), which are close relatives of *Polytomella* algae (Figueroa-Martinez *et al*., 2015). Consequently, *Polytomella* is well situated within the tree of green algae for comparative

studies on carotenoid biosynthesis with model photosynthetic species. The next obvious step is a detailed biochemical characterization of the pigments within *Polytomella* taxa.

Although there has yet to be a transformation system developed for *Polytomella* algae, it is conceivable that they could serve as a biological factory for carotenoid production. *Polytomella* are fast growing (~6 hour doubling time) and can be cultivated at room temperature in a simple, well-defined medium. They also constitute a potentially cost-effective industrial system since, in contrast to photosynthetic algae, their growing conditions do not require a photo-bioreactor and the absence of a cell wall makes cell disruption easy.

More broadly, *Polytomella* algae could yield insights into why the carotenoid pathway (as well as eyespots) are maintained in some heterotrophic algae and lost in others. Although the presence of a carotenoid biosynthetic pathway is considered quite rare among well-studied nonphotosynthetic species, new data from diverse lineages are showing that it is not as rare as once thought. For instance, the nonphotosynthetic apicomplexan parasites *Plasmodium falciparum* and *Toxoplasma gondii* both have red-algal-derived plastids and both can synthesize carotenoids, which appear to have an important metabolic role (neither species has an eyespot) (Nagamune *et al*., 2008; Tonhosolo *et al*., 2009). *Polytomella*, however, appears to be the first example of a nonphotosynthetic genus in which some members have retained the carotenoid pathway whereas others have lost it. If anything, these data support the idea that in nature the evolutionary loss of photosynthesis (at least with respect to *Polytomella* algae) can lead to the loss of carotenoid biosynthesis, rather than the other way around, which some have hypothesized (Inwood *et al*., 2008). It will be interesting to see if other aspects of *Polytomella* nuclear genomic architecture follow a similar pattern to those observed here for the carotenoid pathway. An earlier study on *Polytomella* mitochondrial genomes showed that palindromic genes were lost or retained to various degrees in the four known lineages highlighted here (Smith *et al*., 2013). If we have learned anything about *Polytomella* genomic architecture over the past decade, it is that anything can go.

Acknowledgements

DRS is supported by a Discovery Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

Borza T, Popescu CE, Lee RW. 2005. Multiple metabolic roles for the nonphotosynthetic plastid of the green alga *Prototheca wickerhamii*. *Eukaryotic Cell* **4**: 253–261.

Cazzonelli CI. 2011. Carotenoids in nature: insights from plants and beyond. *Functional Plant Biology* **38**: 833–847.

Cordero BF, Couso I, León R, Rodríguez H, Vargas MÁ. 2011. Enhancement of carotenoids biosynthesis in *Chlamydomonas reinhardtii* by nuclear transformation using a phytoene synthase gene isolated from *Chlorella zofingiensis*. *Applied Microbiology and Biotechnology* **91**: 341–351.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.

Fassett RG, Coombes JS. 2011. Astaxanthin: a potential therapeutic agent in cardiovascular disease. *Marine Drugs* **9**: 447–465.

Figueroa-Martinez F, Nedelcu AM, Smith DR, Reyes-Prieto A. 2015. When the lights go out: the evolutionary fate of free-living colorless green algae. *New Phytologist* **206**: 972–982.

Kreimer G. 2009. The green algal eyespot apparatus: a primordial visual system and more? *Current Genetics* **55**: 19–43.

Ladygin VG, Semenova GA. 2014. Ultrastructural organization and composition of carotenoids of the eyespot in *Chlamydomonas reinhardtii* mutants. *Cell and Tissue Biology* **8**: 267–275.

Lohr M. 2009. Carotenoids. In: Stern D, ed. *The Chlamydomonas Sourcebook, vol. 2: Organellar and Metabolic Processes*. Amsterdam: Academic Press, 799–817.

Lohr M, Im CS, Grossman AR. 2005. Genome-based examination of chlorophyll and carotenoid biosynthesis in *Chlamydomonas reinhardtii*. *Plant Physiology* **138**: 490–515.

Inwood W, Yoshihara C, Zalpuri R, Kim KS, Kustu S. 2008. The ultrastructure of a *Chlamydomonas reinhardtii* mutant strain lacking phytoene synthase resembles that of a colorless alga. *Molecular Plant*, **1**: 925–937.

MacDonald SM, Lee RW. 2015. Validation of *Polytomella piriformis* nomen nudum (Chlamydomonadaceae): a distinct lineage within a genus of Non-Photosynthetic Green Algae. *Journal of Eukaryotic Microbiology*, in press.

Nagamune K, Hicks LM, Fux B, Brossier F, Chini EN, Sibley LD. 2008. Abscisic acid controls calcium-dependent egress and development in *Toxoplasma gondii*. *Nature* **451**: 207–210.

Pombert JF, Blouin NA, Lane C, Boucias D, Keeling PJ. 2014. A lack of parasitic reduction in the obligate parasitic green alga *Helicosporidium*. *PLoS Genetics* **10**: e1004355.

Pringsheim EG. 1955. The genus *Polytomella*. *Journal of Protozoology* **2**: 137–145.

Santabarbara S, Casazza AP, Ali K, Economou CK, Wannathong T, Zito F, Redding KE, Rapparport F, Purton S. 2013. The requirement for carotenoids in the assembly and function of the photosynthetic complexes in *Chlamydomonas reinhardtii*. *Plant Physiology* **161**: 535–546.

Shumskaya M, Wurtzel ET. 2013. The carotenoid biosynthetic pathway: thinking in all dimensions. *Plant Science* **208**: 58–63.

Smith DR, Hua J, Archibald JM, Lee RW. 2013. Palindromic genes in the linear mitochondrial genome of the nonphotosynthetic green alga *Polytomella magna*. *Genome Biology and Evolution* **5**: 1661–1667.

Smith DR, Lee RW. 2014. A plastid without a genome: evidence from the nonphotosynthetic green alga *Polytomella*. *Plant Physiology* **164**: 1812–1816.

Tonhosolo R, Fabio LD, de Rosso VV, Gazarini ML, Matsumura MY, Peres VJ, Merino EF, Carlton JM, Wunderlich G, Mercadante AZ *et al.* 2009. Carotenoid biosynthesis in intraerythrocytic stages of *Plasmodium falciparum*. *Journal of Biological Chemistry* **284**: 9974–9985.

Ye X, Al-Babili S, Klöti A, Zhang J, Lucca P, Beyer P, Potrykus I. 2000. Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* **287**: 303–305.

Supporting Information

Methods S1. Mining carotenoid genes from *Polytomella* transcriptomes and genomes.

Table S1. Carotenoid pathway transcript sequences and their sources for *C. reinhardtii*, *Volvox carteri*, and *Polytomella* algae.

Figure Legends

Figure 1. Retention and loss of the carotenoid biosynthetic pathway in photosynthetic and nonphotosynthetic green algae. (A) The putative carotenoid biosynthetic pathway for *Chlamydomonas reinhardtii* and other green algae based on Lohr *et al*. (2005) and Lohr (2009). Enzymes are numbered and placed within black arrows; products are boxed in white. Abbreviations are as follows: Enzymes: 1-deoxy-D-xylulose-5-phosphate synthase (DXS), 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR), 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase (CMS), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK), 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MCS), 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate synthase (HDS), isopentenyl-/dimethylallyl diphosphate synthase (IDS), isopentenyl-diphosphate delta-isomerase (IDI), geranylgeranyl diphosphate synthase (GGPS), phytoene synthase (PSY), phytoene desaturase (PDS), zeta-carotene isomerase (Z-ISO), zeta-carotene desaturase (ZDS), carotenoid isomerase (CRTISO), lycopene β-cyclase (LCYB), carotene β-ketolase (BKT), carotene β-hydroxylase (CHYB), carotenoid β-hydroxylase cytochrome P450 type (LUT5), zeaxanthin epoxidase (ZEP), violaxanthin de-epoxidase (VDE), lycopene ε-cyclase (LCYE), carotene ε-hydroxylase cytochrome P450 type (LUT1). Products: 1-deoxy-D-xylulose-5-phosphate (DXP), 2-C-methyl-D-erythritol-4-phosphate (MEP), 4-diphosphocytidyl-2-C-methyl-D-erythritol (CDP-ME), 4-diphosphocytidyl-2-C-methyl-D-erythritol-2-phosphate (CDP-ME2P), 2-C-methyl-D-erythritol-2,4-cyclodiphosphate (MECDP), 1-hydroxyl-2-methyl-2-(E)-butenyl-4-diphosphate (HMBPP), dimethylallyl-diphosphate (DMAPP), isopentenyl-diphosphate (IPP), geranylgeranyl-diphosphate (GGPP). **(B)** Tree of green algae based on phylogenic analyses of Smith *et al*. (2013) and Figueroa-Martinez *et al*. (2015). Nonphotosynthetic lineages include all four *Polytomella* taxa and *Helicosporidium* sp. **(C)** The presence or

absence of genes encoding carotenoid biosynthetic enzymes in the transcriptomes and genomes of various photosynthetic and nonphotosynthetic green algae. Enzymes are numbered based on pathway shown in part A. A single green checkmark signifies that a complete transcript and gene was identified in both the transcriptome and genome. Red arrows denote the identification of aberrant sequences within the transcriptome and/or genome, including those with large insertions (>150 nt), poor sequence identity, and premature stop codons (shown with a circle). Sources for transcriptomic and genomic data as well as search methods are listed in Supplementary Table S1. **(D)** Amino acid sequence alignments of carotenoid biosynthetic enzymes (numbering and abbreviations as shown above). Alignments were generated with MUSCLE (Edgar, 2004) implemented through Geneious v8.1.4 (Biomatters, New Zealand) using default settings and 8 iterations. Similarity shading generated using the Blosum62 score matric and a threshold value of 1 in Geneious. The LCYB has more premature stop codons than shown on the figure.

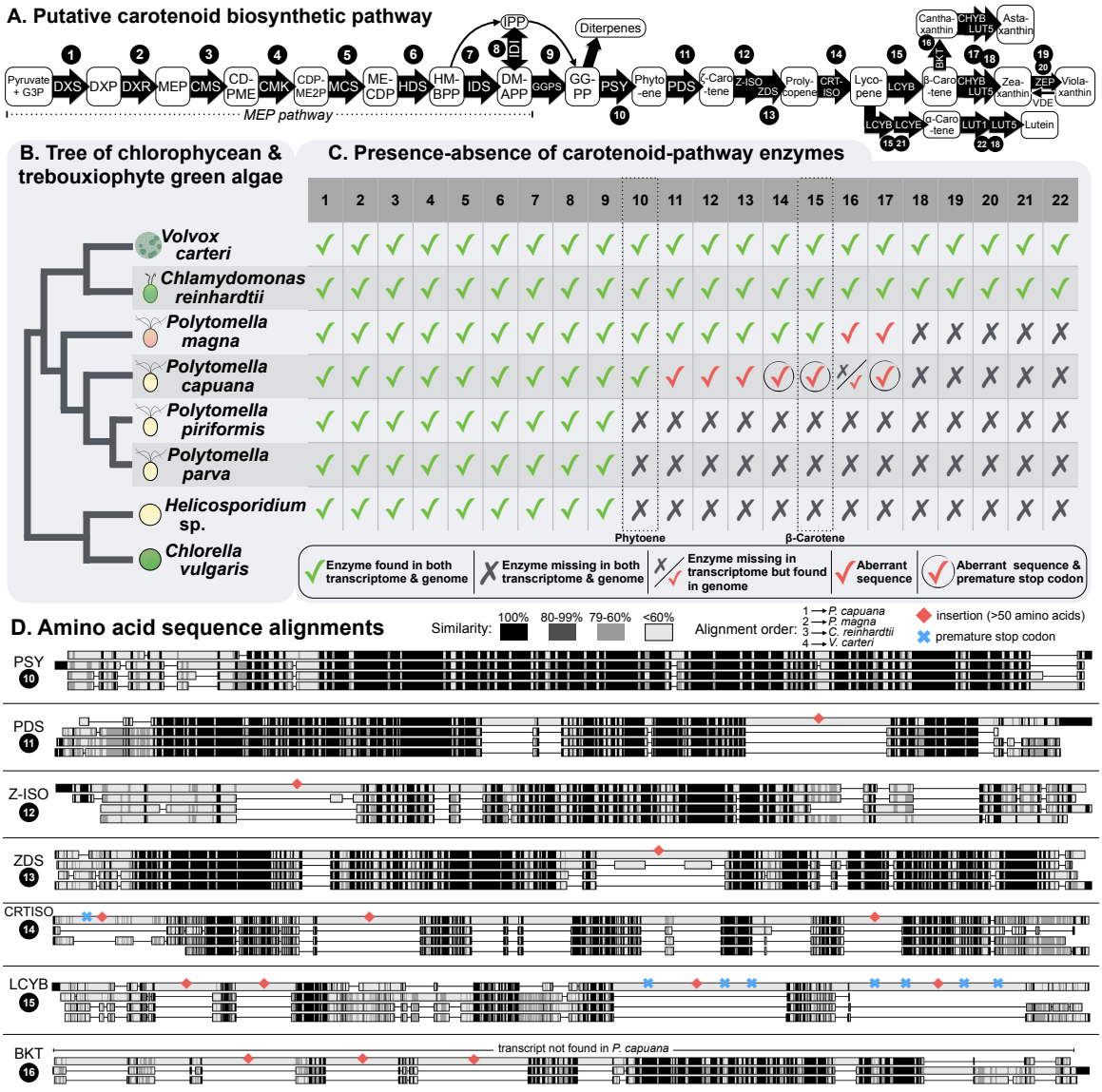


Figure 1: Retention and loss of the carotenoid biosynthetic pathway in photosynthetic and nonphotosynthetic green algae.

Table S1: Carotenoid pathway transcript sequences and their sources for *C. reinhardtii*, *Volvox carteri*, and *Polytomella* algae. Can be found online.

Methods S1. Mining carotenoid genes from *Polytomella* transcriptomes and genomes.

Polytomella piriformis, P. parva, P. capuana, and *P. magna* (Göttingen Culture Collection of Algae, SAG, strains 63-10, 63-3, 63-5, and 63-9, respectively) were grown and harvested as previously described (Smith & Lee, 2014, and references therein). *P.*

parva RNA extraction, library preparation, Illumina sequencing, and *de novo* RNA-Seq assembly were carried out in conjunction with the National Center for Genome Resources (NCGR) following the protocols of the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP) (Smith & Lee, 2014; Keeling *et al*., 2015). For *P. piriformis*, *P. capuana* and *P. magna*, total cellular RNA was isolated using the Qiagen (MD, USA) RNeasy Plant Mini Kit and treated with Qiagen RNase-free DNase followed by RNA-Seq library preparation and Illumina (HiSeq 2500) sequencing (paired end, 2 x 150 cycle run) at the McGill University and Genome Quebec Innovation Centre (MUGQIC). The raw RNA-Seq data were trimmed and clipped with Trimmomatic (Bolger *et al*., 2014) and the normalized reads were assembled with Trinity (Haas *et al.*, 2014) by the MUGQIC Bioinformatics team using their standard parameters. DNA extraction, Illumina sequencing, and *de novo* assemblies of the four *Polytomella* species have been previously described (Smith & Lee, 2014). The nucleotide and deduced amino acid sequences of the *C. reinhardtii* carotenoid genes (Fig. 1a, Supporting Information Table S1) were blasted (Altschul *et al*., 1990) against the *Polytomella* transcriptome and genomes assemblies using TBLASTX and TBLASTN, respectively. Parameters were as follows: Word size: 3; Matrix: BlOSUM62; Max E-value: 1e-10; low complexity filter: off; Gap costs (for TBLASTN): existence 11, extension 1.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Bolger AM, Lohse M, Usadel B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, btu170.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, *et al*. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc* 8, 1494–1512.

Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, *et al.* 2014. The marine microbial eukaryote transcriptome sequencing project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol* 12, e1001889.

Smith DR, Lee RW. 2014. A plastid without a genome: evidence from the nonphotosynthetic green algal genus *Polytomella*. *Plant Phys*. 164, 1812–1819.

Table S1. Carotenoid pathway transcript sequences and their sources for *C. reinhardtii*, *Volvox carteri*, and *Polytomella* algae. See associated Microsoft Excel file.

Curriculum Vitae

Name: Sara Asmail

Post-secondary Education and Degrees:
University of Western Ontario
London, Ontario, Canada
2009-2013 B.Sc.

The University of Western Ontario
London, Ontario, Canada
2013-2015 M.Sc.

Honours and Awards:
Ontario Graduate Scholarship
2013-2015

Dean's Honor Roll
2013

Related Work Experience
Teaching Assistant
The University of Western Ontario
2013-2015

Publications:

Smith, D.R., and Asmail, S.R. 2014. Next-generation sequencing data suggest that certain nonphotosynthetic green plants have lost their plastid genomes. New Phytologist 204: 7-11.

Asmail, S.R., and Smith, D.R. 2015. Retention, erosion, and loss of the carotenoid biosynthetic pathway in the nonphotosynthetic green algal genus *Polytomella*. New Phytologist, in press.